STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 142287

TO: Elizabeth McElwain
Location: REM/2A11/2C18
Art Unit: 1638
Tuesday, January 25, 2005

Case Serial Number: 10/070666

From: Mary Jane Ruhl
Location: Biotech-Chem Library
Remsen 1-A-62
Phone: 571-272-2524

maryjane.ruhl@uspto.gov

Search Notes

Examiner McElwain,

Here are the results for your recent search request.

Please feel free to contact me if you have any questions about these results.

Thank you for using STIC services. We appreciate the opportunity to serve you.

Sincerely,

Mary Jane Ruhl
Technical Information Specialist
STIC
Remsen 1-A-62
Ext. 22524

STIC-Biotech/ChemLib

1442787

From: McElwain, Elizabeth
Sent: Tuesday, January 11, 2005 4:06 PM
To: STIC-Biotech/ChemLib
Subject: sequence search

Please search 10/070,666 - SEQ ID NO: 1 for prior art only

Thank you,
Beth

Elizabeth F. McElwain, Ph.D.
U.S. Patent and Trademark Office
Tech Center 1600, Art Unit 1638
room Remsen 2A11
mailbox Remsen 2C18
571-272-0802
elizabeth.mcelwain@uspto.gov

****
STAFF USE ONLY

Searcher:_____
Searcher Phone: 2-
Date Searcher Picked up:____
Date Completed:_____
Searcher Prep/Rev. Time:___
Online Time:_____

****
Type of Search
NA Sequence: #_____
AA Sequence :#_____
Structure: #_____
Bibliographic:_____
Litigation:_____
Patent Family:_____
Other:_____

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other(Specify):_____

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 19, 2005, 06:21:29 ; Search time 4402 Seconds
(without alignments)
10090.875 Million cell updates/sec

Title: US-10-070-666A-1
Perfect score: 1219
Sequence: 1 agtaagcaaactaaatttaa.....taacaaaaaaaaaaaaaaaa 1219

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		615.4	50.5	659	4	BM399473	BM399473 5009-0-58
2		68.2	5.6	1101	9	CNS00EVL	AL069706 Drosophil
3		66.8	5.5	1101	9	CNS0039G	AL063921 Drosophil
4	c	65.8	5.4	1101	9	CNS0039G	AL063921 Drosophil
5	c	62	5.1	928	9	CNS00DKY	AL071865 Drosophil
6	c	61.8	5.1	1101	9	CNS00EPO	AL069493 Drosophil
7	c	61.4	5.0	1201	9	CNS0167M	AL106396 Drosophil
8		60.4	5.0	886	8	BH177277	BH177277 008_L_22-
9		60.4	5.0	886	9	CNS07JUX	AL614235 T3 end of
10		60.4	5.0	961	9	CNS008HI	AL051882 Drosophil
11	c	60.2	4.9	1101	9	CNS00EVL	AL069706 Drosophil
12	c	60	4.9	1101	9	CNS001FB	AL060732 Drosophil
13		59.8	4.9	945	9	CNS04D0K	AL285149 Tetraodon
14	c	59.2	4.9	1101	9	CNS00KAE	AL077628 Drosophil
15		58.8	4.8	1324	9	AG376784	AG376784 Mus muscu
16		58.4	4.8	1101	9	CNS00Z2U	AL097152 Drosophil
17		58	4.8	1101	9	CNS00BO1	AL057419 Drosophil
18		57.6	4.7	1013	9	CNS06RPQ	AL412260 T7 end of
19	c	57.6	4.7	1101	9	CNS00EO7	AL069440 Drosophil
20	c	56.6	4.6	945	9	CNS04D0K	AL285149 Tetraodon
21		56	4.6	1201	9	CNS0167M	AL106396 Drosophil
22	c	55.6	4.6	1013	9	CNS06RPQ	AL412260 T7 end of
23	c	55.6	4.6	1101	9	CNS0021J	AL061936 Drosophil
24	c	55.4	4.5	1101	9	CNS00BEU	AL056987 Drosophil
25		55.2	4.5	1043	9	CNS0145P	AL103735 Drosophil
26		55.2	4.5	1101	9	CNS00FVE	AL071298 Drosophil
27		55.2	4.5	1101	9	CNS00KAE	AL077628 Drosophil
28		55.2	4.5	1206	5	BQ709824	BQ709824 AGENCOURT
29	c	55	4.5	1101	9	CNS003BD	AL064091 Drosophil
30		54	4.4	928	9	CNS00DKY	AL071865 Drosophil
31		53.8	4.4	839	8	AQ892711	AQ892711 HS_4832_A
32	c	53.8	4.4	900	8	BH135675	BH135675 ENTNT31TF
33	c	53.8	4.4	1101	9	CNS002B4	AL062483 Drosophil
34		53.6	4.4	1101	9	CNS0039L	AL063926 Drosophil
35	c	53.6	4.4	1101	9	CNS008WC	AL052719 Drosophil
36		53.6	4.4	1210	9	CG749728	CG749728 P044-1-C0
37	c	53.4	4.4	914	9	CNS00ZJY	AL097768 Drosophil
38		53.4	4.4	1101	9	CNS00FMC	AL070972 Drosophil
39	c	53.4	4.4	1146	9	CNS021G2	AL176843 Tetraodon
40		53.2	4.4	1101	9	CNS0177R	AL107697 Drosophil
41		53	4.3	1101	9	CNS00LVZ	AL078819 Drosophil
42	c	53	4.3	1168	9	CL078758	CL078758 CH216-153
43		52.8	4.3	477	6	CD135036	CD135036 MG1-0036U
44	c	52.8	4.3	1101	9	CNS00EJ4	AL069257 Drosophil
45	c	52.8	4.3	1101	9	CNS00EQL	AL069526 Drosophil

ALIGNMENTS

RESULT 1
BM399473
LOCUS BM399473 659 bp mRNA linear EST 17-JAN-2002
DEFINITION 5009-0-58-B08.t.1 Chilcoat/Turkewitz cDNA (large fraction)
Tetrahymena thermophila cDNA, mRNA sequence.
ACCESSION BM399473
VERSION BM399473.1 GI:18199526
KEYWORDS EST.
SOURCE Tetrahymena thermophila
ORGANISM Tetrahymena thermophila
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE 1 (bases 1 to 659)
AUTHORS Turkewitz,A.P., Karrer,K.M., Jahn,C., Orias,E., Kirk,K.E.,
Frankel,J. and Klobutcher,L.
TITLE EST from Tetrahymena thermophila, strain CU428.1, growing cells
JOURNAL Unpublished (2002)
COMMENT Contact: Turkewitz AP
Molecular Genetics and Cell Biology
University of Chicago
920 E. 58th Street, Chicago, IL 60637, USA
Tel: 773 702 4374
Fax: 773 702 3172
Email: apturkew@midway.uchicago.edu
Seq primer: T3.
FEATURES Location/Qualifiers
source 1. .659
/organism="Tetrahymena thermophila"
/mol_type="mRNA"
/strain="CU428.1"
/db_xref="taxon:5911"
/clone_lib="Chilcoat/Turkewitz cDNA (large fraction)"
/note="Vector: BlueScript2 SK+; Details on library
preparation can be found in Chilcoat and Turkewitz (2001)
Proc. Natl. Acad. Sci USA, 98: 8709-8713."
ORIGIN

Query Match 50.5%; Score 615.4; DB 4; Length 659;
Best Local Similarity 99.8%; Pred. No. 1e-112;
Matches 616; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 AGTAAGCAAACTAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         43 AGTAAGCAAACTAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAA 102

Qy         61 TTGTTCTTGAAAATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        103 TTGTTCTTGAAAATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTG 162

Qy        121 AATATGACTGCACTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 AATATGACTGCACTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCA 222

Qy        181 ATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 ATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGT 282

Qy        241 AGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        283 AGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAAT 342

Qy        301 CTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 CTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGC 402

Qy        361 CTATCGAAATTGGTTTATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 CTATCGAAATTGGTTTATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTC 462

Qy        421 AAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 AAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTG 522

Qy        481 GTCACTCTATGAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 GTCACTCTATGAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTC 582

Qy        541 CTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 CTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGT 642

Qy        601 TCACAAACAACATTCTA 617
              |||||||||||||||||
Db        643 CCACAAACAACATTCTA 659
```

RESULT 2
CNS00EVL

```
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN
```

Query Match 5.6%; Score 68.2; DB 9; Length 1101;
Best Local Similarity 31.5%; Pred. No. 0.0022;
Matches 192; Conservative 115; Mismatches 299; Indels 3; Gaps 1;

```
Qy        610 ACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCCCCTTCTTAT 669
              |: :|: :::  :   : ::|  :|| :  ||:|  :|      ::::     |: :::|
Db        459 AHTWTMMMMMWWAATWTWWAAAWAAAWTTATWAATWAAAAAWAWWWWATTTTTWWWWWT 518

Qy        670 TTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGAATCTTCC 729
              : ||:   :  : :| : |    |   :|   | | | ::   |: | ||  | |  :
Db        519 WATTWTTTWAWWTWTAWTAAAAAAAAAWATAATTTAAAWWAATAWATTAAWAATTTAWAA 578

Qy        730 TTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAAT 789
              ::    |  | :    | |:|:||| |  |||||      | |  :|     || | |||
Db        579 WWTATATTAATWTATAAATWTWATTAATATAAAAAAATATTTTTTWATAAAATTTTTAAT 638

Qy        790 TGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAAT 849
              ||  |  :|| || | |  :|   |: :|  |  ||    | :|||| :||  |||
Db        639 AATTTAATTAWTTATTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWAAAAAAAA 698

Qy        850 TCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACG 909
              ||| :| :|  | |:    ||  :|  ||    :||||     |:    |:: ||
Db        699 AAAAAAAWATWAAWAATWATAWATAAWTTAAAA---WAATAAAAWAAWAATWAWWATAAT 755

Qy        910 CTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATC 969
              |::  |    :| :| : : |:  :::|||   |   || :||:|: :     | ||
Db        756 ATWWATATATAWTTWTAWWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATAWATA 815

Qy        970 ACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAG 1029
              : |:      |:| |::  : | || : || :   |||  |    |  :|||:    :|
Db        816 WATWAAAWAWAWATAWWATWATATAWWAATAWAWAAAAAATWTAATATWAATWATAAWAA 875

Qy        1030 AATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTT 1089
               ||   |: :  :  :   |::| : |    |:  |  :|  :   |||||    |   :
Db        876 AAAATAWAWTTWWTWTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAAATAAWA 935

Qy        1090 GAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAA 1149
               :: :  | |:   |||  :|: ||  || : ||:: : : ||| | : | :| ||   :
Db        936 WWWTWATATTWTTATTAAAWTWTATWWATTWATTWWAWTWTATAWTTWTATWTATATATW 995

Qy        1150 ATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAA 1209
               :| :  :||   ||  ||| ::  || ||| :||  | | | |   :: | | | |:
Db        996 WTAWTAWATATATTTATTAAWWTATATTTTAWAAAWTAATATATWATAWWWTWATATATAW 1055

Qy        1210 AAAAAAAAA 1218
               |::||  |:
Db        1056 AWWAATTAW 1064
```

RESULT 3
CNS0039G

```
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
```

```
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES                Location/Qualifiers
     source             1. .1101
                        /organism="Drosophila melanogaster"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7227"
                        /clone="BACR08K10"
                        /clone_lib="RPCI-98"
                        /note="end : TET3"
ORIGIN

  Query Match              5.5%;  Score 66.8;  DB 9;  Length 1101;
  Best Local Similarity   18.5%;  Pred. No. 0.0043;
  Matches  126;  Conservative  273;  Mismatches  282;  Indels    1;  Gaps    1;

Qy          12 TAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAATTGTTCTTGAA 71
               :|||:  || ||  ::  : :|    | |: | || |   :|  | :  |  |:   ||
Db         417 WAAAWAAAATAATTWWAAWAWAAAAAATTWWAAAAWAAAAWAWTAWWTTTAWTWAAAAAA 476

Qy          72 AATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTGAATATGACTGC 131
               || |||        || :       :| |::||:::||| |:  | |   || :  |
Db         477 AAAAAATTWTTTTTTTWTTTAWTTWATAWWTTWWWTTAAAWAAAAAAAAAAAAWAAAWAAA 536

Qy         132 ACTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCAATTTGTTTATT 191
               : | |||:|  |:: |:  : |:  |  :       ::|:  :|::::  :|:| |:: :|
Db         537 WATAAATWTWWTWWTTYTTWAAWATAAAMCMAAWYYHTYTTYTYHYYTYWTYTMTWHYHT 596

Qy         192 GATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGTAGGCTTTGAAA 251
                 | | |: |:  :::| |:: ::|:::|:: :::: |: | :::    :  :::  :::
Db         597 MYTHAWAHTTWYHWYHTYAMWHWMTWHWTMWAWWHWTTYTAAYYYYYTCMYYYHYMHWHH 656

Qy         252 ATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAATCTTCAAAGAGA 311
               |: : ||::| :|:  :  : | :  : ::  | :        ::    :::: |:: : :
Db         657 AHAHAAAWWTTHTWWTHAYHWATYHYYYYMYCAMMCMCTHTCHHCYYYYHHYTAHHTHTH 716

Qy         312 TTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGCCTATCGAAATT 371
               ::: :   :::::|  ::      | : ::::: ::  :    |: :   ::::   ||
Db         717 HWYAHYYMWYYMWAYYWMYCTACTYHYHHHHHYHWAYHTTWYAWAHAMWMWHHAHYAAAA 776

Qy         372 GGTTTATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTCA-AAAGTGGTA 430
                :|||:::::  |:::::|  ::::  :|  ::||   : ::: |:||: : : :  |:
Db         777 AWAAWATTHHYHHTTHYMHHTYMYHYYMYTCCYMCTYHCWHYYHTAYTCWTWTHHWMWTW 836

Qy         431 TTTCTCTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTGGTCACTCTAT 490
               |:::::|::::   :: :: |:|: ::  :| |  :          :::   :::::::
Db         837 THWYHHTWWHHTTTHWAWWHTHTWCWWWWHATTWTWATHCWACMTMHWHHWMHMHHHHMA 896

Qy         491 GAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTCCTCTTTGTGG 550
                :|::: :: : :::::|| :::: : ::  ::: : ::: :::::  :: |:||
Db         897 CHAHHHTHMCMCHHHMHCTCHHHHTMYHMTCHWMWHMHWHWHHWWMATWMTTMTTMMMMC 956

Qy         551 TGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGTTCACAAACAA 610
```

```
                  :::|:::::  ::          |   :|  :::  |  :::| ::| :|   :|: |:
Db         957 CMMHHHCHMYHMMHMYMYCCHYYCTCHTHATTHYHYMCTCYHYCTWHTYWTAYWWAWTAH 1016

Qy         611 CATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCCCCTTCTTATT 670
                :|| |:::  :  |: ||:: ::: :  |::||  ::: :  |: |:::::|:: :: :
Db        1017 AMTTATWWWWMHWWAHWATWWWWWWWWATAWWACTCHHWTWYHTHCTWYYHHTYHMWWAW 1076

Qy         671 TTTAAAGTGGAAATTAGACTCC 692
               :  ::: :  |: ::| :|:::
Db        1077 WMAWHWHHMYAHYHWAHHCWYY 1098


RESULT 4
CNS0039G/c
LOCUS       CNS0039G                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES                Location/Qualifiers
     source             1. .1101
                        /organism="Drosophila melanogaster"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:7227"
                        /clone="BACR08K10"
                        /clone_lib="RPCI-98"
                        /note="end : TET3"
ORIGIN

  Query Match              5.4%;  Score 65.8;  DB 9;  Length 1101;
  Best Local Similarity   16.6%;  Pred. No. 0.0068;
  Matches   97;  Conservative  250;  Mismatches  238;  Indels    0;  Gaps    0;

Qy         633 TAACACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTAAAGTGGAAATTAGACTCC 692
               ::: | :|: : :::  ||  || ::    :: |::|:::::::    |  :||   :
Db        1067 DRRWAGDADRWAWDDGAGTWWTATWWWWWWWWATWDTWWDKWWWWATAAKTDTAWTWWRT 1008

Qy         693 ATCTTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTA 752
               |:  ::   :    : : :|||: : ||:::   :  :::  ::| : ::||: ::| :|
Db        1007 AWRADWAGRDRGAGKRDRDAATDADGAGRRDGGRKRKDKKDRKDGDDDKKGGKKKKAAKA 948

Qy         753 TTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCT 812
               : |  ::::: :: :::::  ::::  :::  |: ::: : :  :::|: |: :::: :
Db         947 AKWATKWWDDWDWDKDWKWDGAKDRKADDDDGAGDKDDDGKGKDADDDTDGTKDDDDKDK 888
```

```
Qy        813 TCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCA 872
              : :::: :  : :  :|: ||:::::  |:|: : ::   ::: |::::: : ::::  :
Db        887 WDDWDKAKGTWGDATWAWAATDWWWWGWADADWWTWDAAADDWWADDRWDAWAWKWDDAW 828

Qy        873 TTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTA 932
              :     :: :: ::|  :     :: :::|: ::   ::    |  :  : || | : ::
Db        827 AWGARTADRRDWGDRAGKRGGARKRRDRKRADDKRDAADDRDDAATWTTWTTTTTRDTDD 768

Qy        933 CTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTC 992
               :: |: : :   :  :: ::::: ::| |  : :    :  ::     ::: :  :|
Db        767 WKWKTDTWTRWAADRTWDRDDDDDRDRAGTAGRKWRRTWKRRWKRRDTRWDDADADDTAR 708

Qy        993 TACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAG 1052
              :: ::: ::   |  | : | :  : :: |    :| :|: |::   :|: ::  : :::
Db        707 DDRRRRGDDGADAGKGKKTGRKRRRRDRATWDRTDAWWADAAWWTTTDTDTDDWDKRDRR 648

Qy       1053 ATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAA 1112
              ::  : :::  |   ::::|::|:::|: :     :  :::| : :|:|:: :::: :|:
Db        647 RKGARRRRRTTARAAWDWWTWKAWDWAKWDWKTRADRWDRWAADTWTDARKADRDWAKAR 588

Qy       1113 ATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTA 1172
              |:: : :    |  :: : |  :   | :| :| :||:: :  :||   |: ||:  |:
Db        587 AWRARRDRARAARADRRWTTKGKTTTATWTTWAARAAWWAWWAWATTTATWTTTWTTTWT 528

Qy       1173 TTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAA 1217
              ||  ||||| :| | | ::|| :   : |: |||:|||||||:||
Db        527 TTTTTTTTTTWTTTAAWWWAAWWTATWAAWTAAAWAAAAAAAWAA 483


RESULT 5
CNS00DKY/c
LOCUS       CNS00DKY                928 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 928)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .928
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR27A24"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN
```

```
  Query Match              5.1%;  Score 62;  DB 9;  Length 928;
  Best Local Similarity   31.1%;  Pred. No. 0.039;
  Matches  122;  Conservative   86;  Mismatches  184;  Indels    0;  Gaps    0;

Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
              |||  |  ||||  |||  |||    |||  || ||  |    |  |       ||  |  |
Db        922 AAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 863

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGG 946
              ||::     ||  : :::| :: :   :::: :::  |  :::::::::::: |   |
Db        862 AAWWYTCMAAAMAMAMMMCVMMMMMMMMMMMMMMMMTTMWWHHHWHHHHHHWTTTTTTTW 803

Qy        947 TATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAA 1006
              | | |  | ||  : : :: :: : ||:::   :::::::  |   : :  : :|   ::
Db        802 TTTTTTTTTTTTTWWWWWWWWWWWWTTTWWWWTWWWWWWWWWWTNNNWNMCHWAWATWWWW 743

Qy       1007 AGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACC 1066
              :  |  | : ||: :   | :|  ||  ||| ||    ||  ||    ||  |:  :  |
Db        742 WWATATTTWTATWTWAWATAWAAAAAAAAAAAAAAWTAAATTTTTWTTTTAWTTWTAAAA 683

Qy       1067 TATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATT 1126
              | ||||   :   | |    |:|  : ||| | :|:| :|:  :|:||      |: : |
Db        682 TTTTTTATTWTTTTTTWTTTTWTTTWTATATAAWAWTAWAWAAWAWATATTATATWAWAT 623

Qy       1127 AGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCT 1186
              | | ::  | : :  |  |  ::   : |||     : | :| |||  | |:   :|
Db        622 ATTTWWTATTWAWTATAAAATWWTAWWTTTAAWTAAWATAWAATTAAAAWTWAAAWAAAA 563

Qy       1187 GAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1218
              | |:|| |   || : :|||||::|||||:|
Db        562 AAAAWAAAATAWTTTWTWAAAAAWWAAAAAWA 531


RESULT 6
CNS00EPO/c
LOCUS       CNS00EPO                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29P01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069493
VERSION     AL069493.1  GI:4949636
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
```

```
                     /db_xref="taxon:7227"
                     /clone="BACR29P01"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             5.1%;  Score 61.8;  DB 9;  Length 1101;
  Best Local Similarity   34.8%;  Pred. No. 0.043;
  Matches  207;  Conservative   85;  Mismatches  300;  Indels    2;  Gaps    1;

Qy        600 TTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTC 659
              :: | ||  ||  || || ||:    |  ::| |     :|||     || : :|  |
Db       1081 DWTADAAGRAATTTTDTATAGKTGAGAATWRTGTKTWTTKATTTTTTTTTTDWRTGTTAT 1022

Qy        660 CCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAA 719
                | |||:|::  :|::  :|:   || |   | || : || |::: :  ::|||  :
Db       1021 RRRTGTTTAKTKWATRARWWTWADGATGAGTTDGTTTTTRTTTRTKDKAWAWWWTTTTKW 962

Qy        720 GGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATT 779
              |: |  :  ||   |   : |    ::||  ||| |   ||:| || :|: :  |||:
Db        961 TGWTTWGWTWTTAKDTRTTDWTWTTADKTA--ATTGAGTGAAWAKTTRWAWTWWATAAKA 904

Qy        780 CTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAAT 839
              | |   : :|:  |:   :|:| |: |: | ||  ||     : |    : ||| ||||
Db        903 TTRTAGARTKTKTGTRRATRTDTTTDKAKAGTTTTGATRGAGAKAGATTTWTATGTAAAT 844

Qy        840 GAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGC 899
              | :|  :: |  ||  || ||  | |:|   : | |:  ||  |   | :   : ::
Db        843 AATDTAGWWATAAAATTAAAAARAAWTWTTTTWATATWAWAAATTRTTAAWAAARWDRTA 784

Qy        900 AGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAG 959
              | |:| ||   ||   |  | |:  | | |          |  :  | : |  | : |
Db        783 AAAWAATATAATTRTTAAAAAAWTATTTTTTATAAATAATTTWAWTWTTWWTTTTTWGAT 724

Qy        960 ACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATA 1019
              | : |: |  | : :|    | | |     |: |:   ||      |   | | |:||:
Db        723 AAWWAWAAAAAAKAWTAKTRARATAAATTATWATWATTATATATWATTTTTATTTTTATW 664

Qy       1020 ATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAA 1079
              |:|  : |   | |||| |  :  |     | |     || :|  |   | |||| :|:|
Db        663 AWTAADWATTTAATAAATATAWTAATTTATTAATAAAATAWATWAAAAAATTTTARADAT 604

Qy       1080 TCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACA 1139
              |      :     |: | | |:| |||:|  ||||   ::|| ||||| | :|  |||
Db        603 TAAATTAWAAWTTAWAATAAAWAATTAWTGAAAATTTTWWTTATATTAWTTWTTKTAAKT 544

Qy       1140 ATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAA 1193
              |   |:  |:||   ||    | : |||  :| ||  ::||| ||   |   ||
Db        543 AAAATRTTAWATTAATTTATWTAWATTTTWWTTTTTTWWTTTTATTAAAATNAA 490


RESULT 7
CNS0167M/c
LOCUS       CNS0167M                1201 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN15M24 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106396
VERSION     AL106396.1  GI:5621701
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN15M24"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
ORIGIN

  Query Match             5.0%;  Score 61.4;  DB 9;  Length 1201;
  Best Local Similarity   39.1%;  Pred. No. 0.051;
  Matches  156;  Conservative   40;  Mismatches  203;  Indels    0;  Gaps    0;

Qy        821 TGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGA 880
              | |   |||   |  |||  |:|  | |::  | | |: |:  | :::|   |  | |
Db       1156 TATAWNAAAAANTATAAAAAAWAAAATAWWATATAAAWAAWAAAWWWWATATTAAATAWA 1097

Qy        881 ACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTAT 940
              | || | : |     ||  | ||| ::   :|   |    : ||      :  |:|| |
Db       1096 ATATAANWAAAAAAAAAAAAAAAAAWWWTTTHTANAATATTTWTTTWNTATAWAATWTTTT 1037

Qy        941 GGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATT 1000
              |  | | | |||: : | : || | | ||||    |||||:|  ||  :| |  ||
Db       1036 TTTTTTTTTTTATAWAWAAAWAAAAAAAAATTTTAAAAATAAAWTAATTATWAAAATTT 977

Qy       1001 ACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGA 1060
              |||  |  : | :|  ||    || ||  :|| |     :   ||   ||: || |
Db        976 TTAAAAATTTTTWATWWTTTTTWAAAAAAAAAAWAATATWAAAWTTTTTTTTATWTATAA 917

Qy       1061 AGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATAT 1120
              |  :  | ||||:   |||    |  :|| |||||: | |||||       ||   | |
Db        916 AWAWAWTTTTTTYWTTAAAAAAAAAAWTTAAAAATWTAAAAATTATAAAATAAAAWAAAA 857

Qy       1121 TTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTT 1180
              |  |  || | :|| ||:| |:   |   : || |         |   |||   ||:
Db        856 AAAAAAAAAAAAAWWAAAAAWTTTWATTATAAWATTTTTWAAAAAAAAAAAATTTAATTY 797

Qy       1181 TAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219
              |  |   | | || |     |: :::|||||| :: |||
Db        796 TTTTTNAAAATAAAAAAWAAAWAWWWAAAAAATWWTAAA 758


RESULT 8
BH177277
LOCUS       BH177277                886 bp    DNA     linear   GSS 19-OCT-2001
DEFINITION  008_L_22-rev SmBAC1 Schistosoma mansoni genomic clone 008L22 5',
            genomic survey sequence.
ACCESSION   BH177277
VERSION     BH177277.1  GI:16275873
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 886)
  AUTHORS   Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
            Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
COMMENT     Other_GSSs: 008_L_22-21
            Contact: Pierce RJ
            INSERM U 167
```

Institut Pasteur de Lille
1 rue du Professeur A. Calmette, 59019-Lille, France
Tel: (33) (0)3 20877783
Fax: (33) (0)3 20877888
Email: Raymond.Pierce@pasteur-lille.fr
CNS sequencing ID=DG0AA008DF11BP1
Plate: 008 row: L column: 22
Seq primer: M13 reverse primer
Class: BAC ends
High quality sequence stop: 886.

```
FEATURES             Location/Qualifiers
     source          1. .886
                     /organism="Schistosoma mansoni"
                     /mol_type="genomic DNA"
                     /strain="Puerto-Rican"
                     /db_xref="taxon:6183"
                     /clone="008L22"
                     /sex="mixed"
                     /dev_stage="cercariae"
                     /lab_host="Biomphalaria glabrata"
                     /clone_lib="SmBAC1"
                     /note="Vector: pBeloBAC 11; Site_1: Hind III; Partially
                     Hind III digested and size-selected S. mansoni  cercarial
                     DNA was ligated into Hind III digested pBeloBAC 11 vector
                     and used to transform E. coli DH10B. The complete library
                     contains 23808 clones from 4 independent
                     sizing-ligation-transformations.  Average insert size
                     ranges from 70-127 kb and genome coverage is 7.9-fold."
ORIGIN

  Query Match             5.0%;  Score 60.4;  DB 8;  Length 886;
  Best Local Similarity   39.9%;  Pred. No. 0.08;
  Matches  157;  Conservative   47;  Mismatches  183;  Indels    6;  Gaps    1;

Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
              ||||  :    | :  | || :| ||  | |      ||:   ||:::|||: |    | |
Db        416 AAATAMATYASMSNATATAVATTTAAATAATTTTATMTANTTMMMATTWATTATAAAMAT 475

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGG 946
              ::|: :  :||  :  ::: |      :::  |:: ||::| | | :| |  |:
Db        476 WWTMAMAAMAAMTMMTMMMMATYTMMMMHWAMAMMAATWMTTTTTTMTAAMAAWTTTATT 535

Qy        947 TATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAA 1006
              ||:  || :| | | : |  ||   |:  |   |  |: | |   ||   :       |
Db        536 TAWAAAAAWTAAWATWAATAATTTTTWAATTATAATAWATTWTAATTATTTWAATTATTA 595

Qy       1007 AGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACC 1066
              :  :  :  :|||||  | || ||  ||  :  |      |  | | :  | |    |
Db        596 WTTWATWAAAWTAATTTTTTAAAAAAAAATAWTTT------TATAAAAWTTTTATTAATW 649

Qy       1067 TATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATT 1126
              ||:|||  |||||| | |  ||  |||:| || || ||| |||:  |    |||| |
Db        650 TAWTTTAAAAAAATTTTAATTTATAAAWTTAAAAATTTTTTTTTWTTTAATAATTTAAAA 709

Qy       1127 AGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCT 1186
              | ||| | ||| |||| ||  |||    | ||  | |  :|  |   :| ||   |:  :
Db        710 ATTAAAAATAAAAATTTTAAAAAAAAAWTAATAATNTAAWTTTTATAWAATTAAAAWAAW 769

Qy       1187 GAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219
               |:||||   : | |  ||    ||||||||||
Db        770 AARAAAAWTAWTTAATAAATTTTAAAAAAAAAA 802


RESULT 9
CNS07JUX
LOCUS       CNS07JUX                886 bp    DNA     linear   GSS 30-NOV-2001
DEFINITION  T3 end of clone 008DF11 of library SmBAC1 from strain Puerto-Rican
            of Schistosoma mansoni, genomic survey sequence.
ACCESSION   AL614235
VERSION     AL614235.1  GI:16027459
KEYWORDS    GSS.
SOURCE      Schistosoma mansoni
  ORGANISM  Schistosoma mansoni
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 886)
  AUTHORS   Le Paslier,M.C., Pierce,R.J., Merlin,F., Hirai,H., Wu,W.,
            Williams,D.L., Johnston,D., LoVerde,P.T. and Le Paslier,D.
  TITLE     Construction and characterization of a Schistosoma mansoni
            bacterial artificial chromosome library
  JOURNAL   Genomics 65 (2), 87-94 (2000)
  MEDLINE   20247247
   PUBMED   10783255
REFERENCE   2  (bases 1 to 886)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Partially Hind III digested and size-selected S. mansoni cercarial
            DNA was ligated into Hind III digested pBeloBAC 11 vector and used
            to transform E. coli DH10B. The complete library contains 23808
            clones from 4 independent sizing-ligation-transformations. Average
            insert size ranges from 70-127 kb and genome coverage is 7.9-fold.
FEATURES             Location/Qualifiers
     source          1. .886
                     /organism="Schistosoma mansoni"
                     /mol_type="genomic DNA"
                     /strain="Puerto-Rican"
                     /db_xref="taxon:6183"
                     /clone="008DF11"
                     /clone_lib="SmBAC1"
                     /note="end : T3"
ORIGIN

  Query Match             5.0%;  Score 60.4;  DB 9;  Length 886;
  Best Local Similarity   39.9%;  Pred. No. 0.08;
  Matches  157;  Conservative   47;  Mismatches  183;  Indels    6;  Gaps    1;

Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
              ||||  :    | :  | || :| ||  | |      ||:   ||:::|||: |    | |
Db        416 AAATAMATYASMSNATATAVATTTAAATAATTTTATMTANTTMMMATTWATTATAAAMAT 475

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGG 946
              ::|: :  :||  :  ::: |      :::  |:: ||::| | | :| |  |:
Db        476 WWTMAMAAMAAMTMMTMMMMATYTMMMMHWAMAMMAATWMTTTTTTMTAAMAAWTTTATT 535

Qy        947 TATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAA 1006
              ||:  || :| | | : |  ||   |:  |   |  |: | |   ||   :       |
Db        536 TAWAAAAAWTAAWATWAATAATTTTTWAATTATAATAWATTWTAATTATTTWAATTATTA 595

Qy       1007 AGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACC 1066
              :  :  :  :|||||  | || ||  ||  :  |      |  | | :  | |    |
Db        596 WTTWATWAAAWTAATTTTTTAAAAAAAAATAWTTT------TATAAAAWTTTTATTAATW 649

Qy       1067 TATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATT 1126
              ||:|||  |||||| | |  ||  |||:| || || ||| |||:  |    |||| |
Db        650 TAWTTTAAAAAAATTTTAATTTATAAAWTTAAAAATTTTTTTTTWTTTAATAATTTAAAA 709

Qy       1127 AGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCT 1186
              | ||| | ||| |||| ||  |||    | ||  | |  :|  |   :| ||   |:  :
Db        710 ATTAAAAATAAAAATTTTAAAAAAAAAWTAATAATNTAAWTTTTATAWAATTAAAAWAAW 769

Qy       1187 GAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219
               |:||||   : | |  ||    ||||||||||
Db        770 AARAAAAWTAWTTAATAAATTTTAAAAAAAAAA 802


RESULT 10
CNS008HI
LOCUS       CNS008HI                961 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR17O01 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL051882
VERSION     AL051882.1  GI:4933734
```

```
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 961)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .961
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR17O01"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             5.0%;  Score 60.4;  DB 9;  Length 961;
  Best Local Similarity   39.4%;  Pred. No. 0.081;
  Matches  224;  Conservative   46;  Mismatches  290;  Indels    8;  Gaps    1;

Qy        647 ATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTA 706
              :|| |  |  ||    || || |||| ||||  |   ||  | |     ||  |  |||
Db         13 DTTTTTTTTTTTTATNTTTTTTTTTTTAAAAGAAGCKKTTTCAATGACCGTTGNCGGCTTT 72

Qy        707 TTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTT 766
              ||:| | ||| :|    |     || | ||  : |     |||::|||  | | :   |
Db         73 TTWTAATTTTWWATATTTAACATTTTCGTTAAWTTTTAATTTAWWATTTTATTTWTTWTG 132

Qy        767 CTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGG 826
              :| |    ||| ||  |  :||  ||  :  ||| ||   |     ||      |
Db        133 TCWTCTWTATATTATTAATWTWTTTTTTTKWTTTTTTTTTTTTTAACTTACATACATATA 192

Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
              | |  |  ||||  ||   | ||  | |  ||   |   ||    ||||||  |
Db        193 TATTTCTTTAAATAGAAAATACTTTTATAAAATTTTTTTTTTATAAATTTTTAAATTTATT 252

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGG 946
              | ||  |      | |:|    | |          || | | ||||||||||:  |
Db        253 ATTATTTTTTTTATTATATAYTTKTWTAT--------ATTTTTTTWTACTTATTATKAATAT 304

Qy        947 TATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAA 1006
              |||  ||  | : | : : :  || : |||||    |:||||  ||:      :|:
Db        305 TATATATTTWCATWTWTWTTATTWTTTTTTATTTTWAATTTTTTWTATTKTWTWCTTATT 364

Qy       1007 AGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACC 1066
               : :  | | |:| |:  |  |     ||          :|   |:  :|  |     :
Db        365 TKWWTTTWTTAWATTKTTTWTAATYATATTTGTTAATWATAATWTTTWTKTTATTTTWWT 424

Qy       1067 TATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATT 1126
              | ||||| ||:|||| ||  :| :  || : |: :||: ||||      ||||:||  :
Db        425 TTTTTTTTAAWAATCATACAAWTTWTTATTWTTWTWTTKTTTTTTTTATTATATKTTTAY 484
```

```
Qy       1127 AGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCT 1186
              .| | :| | || ||   | : | |  |  | | ||   |: :|: |:  |  | ||
Db        485 TTTTTTTTWATAATTTTTATATATKTTTTATTAATATTTTATATYTWTWGTWAATTTTTT 544

Qy       1187 GAGAAAACAGTCTTAACAAAAAAAAAAA 1214
                :| :   |||  ||| | |||||
Db        545 TTATTWATWTATTTATAAAATAGAAAAA 572
```

```
RESULT 11
CNS00EVL/c
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
ORIGIN

  Query Match             4.9%;  Score 60.2;  DB 9;  Length 1101;
  Best Local Similarity   33.0%;  Pred. No. 0.089;
  Matches  181;  Conservative   92;  Mismatches  275;  Indels    1;  Gaps    1;

Qy        666 TTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGAATC 725
              | :|||:|||| |  |::|||       || |:   |  :   ||| :|| : |:   |:
Db       1033 TAWTTTWTAAAATATAWWTTAATAAATATATWTAWTAWWATATATAWATAWAAWTATAWA 974

Qy        726 TTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCT 785
              :|       | |     | | || |:   | :|::: :|   | |  |  || |:| |
Db        973 WTWWAATWAATWWATAWAWTTTAATAAWAATATWAWWWTWTTATTTTTTTTTTTTTWTTT 914

Qy        786 GAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATG 845
              ||: :||  |  ::| :  | :  :  : :| || ||    || || :| || |:||
Db        913 WTATWTWTTTATATWWTWWAAAAWAWWAAWTWTATTTTTWTTATWATTWATATTAWATT 854

Qy        846 AAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAAC 905
              |:   |  :: |   : |:: :|:|:|  :| :| | :||  |  :   :  :
Db        853 TTTTWTWTATTWWTATATWATWWTATWTWTWTTTWATWT-ATWTATTTATWTWTWTTWTA 795
```

```
Qy        906 TACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAA 965
Db        794 TTWTATWTATAWWWTWWATWWWTAWAAWTATATATWWATATTATWWTWATTWTTWTTTTA 735

Qy        966 CATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCT 1025
Db        734 TTWTTTTAAWTTATWTATWATTWTTWATWTTTTTTTTTTTTTTTWTATTTWAAATTATTT 675

Qy       1026 GAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCAC 1085
Db        674 AATTAWWWAATWAWTTATTTAATAAWTAATTAAATTATTAAAAATTTTATWAAAAAATAT 615

Qy       1086 CTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTA 1145
Db        614 TTTTTTATATTAATWAWATTTATAWATTAATATAWWTTWTAAATTWTTAATWTATTWWTT 555

Qy       1146 GGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAA 1205
Db        554 TAAATTATWTTTTTTTTTTAWTAWAWWTWAAAWAATWAWWWWWAAAAATWWWWTWTTTTTW 495

Qy       1206 AAAAAAAAA 1214
Db        494 ATTWATAAW 486
```

RESULT 12
CNS001FB/c
LOCUS CNS001FB 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR04A23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL060732
VERSION AL060732.1 GI:4939397
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR04A23"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN

Query Match 4.9%; Score 60; DB 9; Length 1101;
Best Local Similarity 38.7%; Pred. No. 0.097;

Matches 152; Conservative 58; Mismatches 180; Indels 3; Gaps 2;

```
Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
Db       1072 WWWAAATWAATAAAATATWAAATWATAWATAATAAAAAAAATATAAATWATTAAAAATAA 1013

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGG 946
Db       1012 AATW--TAAAAAWATWAAAWAAAAAATTWAATAATTAWTATATTNAAAAAAAAWAAAAAA 955

Qy        947 TATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAA 1006
Db        954 AAATTAATAAWWAAAAAAAAAAWATAATAATATATAAAWWTAATWTAATWNATTWAAAAA 895

Qy       1007 AGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACC 1066
Db        894 ATAATAWAATAWWAAATTAAWAAAAATAWATWAAAAAWNTAWWTATAAW-TTWAAAAATW 836

Qy       1067 TATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATT 1126
Db        835 TNTWAWWWAAAAATWWTTTTTTAATATATATWTAATWTAATWTWWAAAAAAAATAWAAWA 776

Qy       1127 AGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCT 1186
Db        775 WTAWWTWAWWWNAWAAAWAAATAATATATAWNAAWAAAAATTWAWWANATWWWTWWWTAAAA 716

Qy       1187 GAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219
Db        715 TAAAAAAAAAATAWATAAATAAWAAAAAAAAAA 683
```

RESULT 13
CNS04D0K
LOCUS CNS04D0K 945 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone 101H21 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL285149
VERSION AL285149.1 GI:8023560
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143
REFERENCE 3 (bases 1 to 945)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at

```
          http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="101H21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG101CD11LP1~end : T7"
ORIGIN

  Query Match             4.9%;  Score 59.8;  DB 9;  Length 945;
  Best Local Similarity   40.1%;  Pred. No. 0.11;
  Matches  174;  Conservative   48;  Mismatches  209;  Indels    3;  Gaps    2;

Qy        789 TTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAA 848
              |:  ||  |  |     :: |  :: |:|| || ||  || | || |||:| ||| . |
Db        433 TWNTTTTTTTTTANTANYWTTAWWWTTWTTTTTTTTATAATTAATAAAAWTTAAANWATT 492

Qy        849 TTCGAAAGAAGAAT--CACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACT 906
              :|  ||| |: |||    |  | |  :::|||  || |  |:|||     |     |: :
Db        493 WTWAAAAWAWAAATAAAAAATAAWAWWWTTTTTAAAAAAAWATAATAAWATTWTWTAWAW 552

Qy        907 ACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAAC 966
              :   ||:  |     ||||   :     ||| |    |  ||   |: | ||   | ||
Db        553 WATTTTWATATTTTTTATTTAAWAATAATATAAAAAATTTTAAAAAWAAATATTTTTAAW 612

Qy        967 ATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATA-ATTGCT 1025
              |  |  : |   ||||||   ||| || | : :|   |::  :  :   ::: |:
Db        613 AATAAAWATAAAAATAAATATATTTATAWAAWAWATAWAWWWWWTAWWAAWWWTAWAATA 672

Qy       1026 GAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCAC 1085
              || || |||| :|    ||   ||| ||  |  :|  :   |::|:  ||:|     |
Db        673 TAAAAAATAAAAWAWAATAAAAATAAAATAAAAAWAAAWWAAAWWTWATAAWANAAAAAA 732

Qy       1086 CTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTA 1145
              |  | |  |: |||| ||  ||||||  | | | :||:: |||    || |||| :
Db        733 AAATATATAATAWAAAATATAAATTAAATATAAAATAWATWWATAAATAAAAAAATTTWT 792

Qy       1146 GGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAA 1205
              |||     :        ||| :||||| : |  ||  |   |:   : |  |||  |
Db        793 AATAATAAAAAWATTAAAWATTATWTATTANWATAAAANTTATAAWNTAWNTAATAAATA 852

Qy       1206 AAAAAAAAAAAAAA 1219
              ||| : ||||||||
Db        853 AAATWTAAAAAAAA 866


RESULT 14
CNS00KAE/c
LOCUS       CNS00KAE               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR39P05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077628
VERSION     AL077628.1  GI:4956903
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACR39P05"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
ORIGIN

  Query Match             4.9%;  Score 59.2;  DB 9;  Length 1101;
  Best Local Similarity   38.1%;  Pred. No. 0.14;
  Matches  216;  Conservative   20;  Mismatches  330;  Indels    1;  Gaps    1;

Qy        654 TAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAA 713
              | :|:    :| |::||| |||: |   | |:    |  || |:    | |||  || |
Db        579 TTWTWTTWTWTTTWWTTTATAAWTTTNTATTWTWTATTAATTTWTTAATTTTAAAATTTA 520

Qy        714 TTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTA-TAT 772
              |||      :| ||  |    ||   :   ||||||||    : : ||| |  || |||
Db        519 TTTTNTATTWTWTTWTTATWTTTTATTWATATATTATTAAATTWTWTAAAATAATATTAT 460

Qy        773 CGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCA 832
              ||   | |  ||| |||    .||| ||     ||   || |  ||    || | |
Db        459 TTATTTTAAATATATATTWATTTTATATTTATTATTATATTTATTTTTATTTWATTATTA 400

Qy        833 TGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGC 892
              |    | | |  |||||       | ||  |   |    :||||  |: |   |||
Db        399 TTTTTAATATAATAAATTTTTTNATATAAATAWAATTTTTWTTTTWAAWTAATTAATTTA 340

Qy        893 TGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTA 952
              |   |   | | ||    ||  |  |   || | | || | || ||      | | ||
Db        339 TTATAATTTATAATATATATTATATTATTATTTTTTTATAATAATAATWCWCTATNTNTA 280

Qy        953 ATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCG 1012
              ||||  | |    ||   | | :     |  |       | | || |   |
Db        279 NTATTTNANTNTNTATATNTNTNWNTATATATATATNANWNANANATNANANTATATNTA 220

Qy       1013 TGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTT 1072
              | | : : |  ||  | | |  |    |   |  |    |  |       || |
Db        219 TATNWNWNTCWCTATANANTANANANTNTTATATATNTNANTNTNTATNTANTATANTNA 160

Qy       1073 TGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAAT 1132
              |  |  ||   :   |   | ||  ||| | | |   ||       | | |  ||
Db        159 TATATNATAANWNTNTNTNANATNTATATANTNANANNAATATATATATATCTATATAGA 100

Qy       1133 ACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAA 1192
              |    || ||   :   | |     || |   |||   | | ||| |   |    |
Db         99 NAGANTTTTTATATANWNANTNTANANANTTTATANTNTATANATATNTAAATTATATAT 40

Qy       1193 ACAGTCTTAACAAAAAAAAAAAAAAAA 1219
              |    | || ||||||| || ||||:
Db         39 AATAATTAAATAAAAAAATAATAAAAW 13


RESULT 15
AG376784
LOCUS       AG376784               1324 bp    DNA     linear   GSS 03-JUN-2004
DEFINITION  Mus musculus molossinus DNA, clone:MSMg01-184I09.T7, genomic survey
            sequence.
ACCESSION   AG376784
VERSION     AG376784.1  GI:47987989
KEYWORDS    GSS.
```

```
SOURCE      Mus musculus molossinus
  ORGANISM  Mus musculus molossinus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     BAC end Sequences of Library MSMg01
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1324)
  AUTHORS   Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the mouse BAC library MSMg01. For BAC
            library availability, please contact Kuniya Abe (abe@rtc.riken.jp).
            Tsukuba Institude, Bio Resource Center,
            The Institute of Physical and Chemical Research (RIKEN) 3-1-1
            Koyadai, Tsukuba, 305-0074 Japan
            phone: 81-298-36-9189, fax: 81-298-36-9199
            e-mail: abe@rtc.riken.jp
            PRIMERS
            Sequencing : T7
            LIBRARY
            Vector     : pBACe3.6
            R.Site 1   : EcoRI
            R.Site 2   : EcoRI.
FEATURES             Location/Qualifiers
     source          1..1324
                     /organism="Mus musculus molossinus"
                     /mol_type="genomic DNA"
                     /sub_species="molossinus"
                     /db_xref="taxon:57486"
                     /clone="MSMg01-184I09.T7"
                     /sex="male"
                     /tissue_type="mixture of kidney and spleen"
                     /clone_lib="MSMg01 Mouse Male BAC Library"
ORIGIN

  Query Match             4.8%;  Score 58.8;  DB 9;  Length 1324;
  Best Local Similarity   44.6%;  Pred. No. 0.17;
  Matches  258;  Conservative    0;  Mismatches  316;  Indels    5;  Gaps    1;

Qy        622 ACGAAGATATCTAACACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGA 681
Db        472 AGGNAANAATNTNNNAAAAAAAAAAAAAGGGGGGTTTTNTTTTTTTTTTTTTTTTTTTTT 531

Qy        682 AATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACT 741
Db        532 TTTTATAATATATTTTTTATTTTTTTTTTTTTTTTTATTTATTTTTTTTTTTTTATTAAATT 591

Qy        742 GGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTT 801
Db        592 ATTTTTTTTATTTTTATTTAATTTATTAATTTTTTATTATTTTTTTTATTAATTTTTTTT 651

Qy        802 TCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAA 861
Db        652 TTTATTTTTTTTTTATTATTATTTTTTTTTATTTTTTAATTTTTTTAAATTATAATTTATAT 711

Qy        862 TCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACA 921
Db        712 TTTTATTATTTTTTTTTTATAATATAATTTTTTTATTTATATTTTATATTATTTTTTTAT 771

Qy        922 TATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCAT 981
Db        772 TTTTTTTTTAATTTTTATTATTTTATATTTATATTTATTTTTATATTTAATTAATTATTT 831

Qy        982 AAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGT 1041
Db        832 TAATTTTTAATTTTATTTTTATAATTTTTTTTATTTTATATTTTTATNATTAAATTAATT 891

Qy       1042 GGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAA 1101
Db        892 -----TTTTATATTTATTTATTTTTTATTTTTTAATTTATTTTATTTATATTTGTTTTTT 946

Qy       1102 ATTTATTTTAAATGCATATTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGT 1161
Db        947 TTTTTTATATTTATATTATTTATTTTATTTTTATTAATTTTTTATTATATTTTTTTTTTT 1006

Qy       1162 TTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTT 1200
Db       1007 TTTATTTTATATTTTTTTTTATTATTATATTATNTTTTT 1045

Search completed: January 19, 2005, 16:58:23
Job time : 4409 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 19, 2005, 14:00:14 ; Search time 757 Seconds
(without alignments)
9252.626 Million cell updates/sec

Title: US-10-070-666A-1
Perfect score: 1219
Sequence: 1 agtaagcaaactaaatttaa.....taacaaaaaaaaaaaaaaaa 1219

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		59	4.8	8056	18	US-10-473-126-386	Sequence 386, App
2	c	58	4.8	6103	15	US-10-311-455-1663	Sequence 1663, Ap
3	c	57.4	4.7	8056	18	US-10-473-126-240	Sequence 240, App
4	c	54.6	4.5	8056	18	US-10-473-126-386	Sequence 386, App
5		53.2	4.4	1501	18	US-10-473-126-328	Sequence 328, App
6		52	4.3	640681	9	US-09-790-988-1	Sequence 1, Appli
7	c	51.8	4.2	424	9	US-09-960-352-11218	Sequence 11218, A
8	c	51.8	4.2	7319	15	US-10-311-455-2017	Sequence 2017, Ap
9	c	51.4	4.2	5728	15	US-10-311-455-74	Sequence 74, Appl
10	c	51.2	4.2	5750	15	US-10-311-455-1982	Sequence 1982, Ap
11	c	51.2	4.2	5750	16	US-10-221-714A-434	Sequence 434, App
12	c	51.2	4.2	3673778	15	US-10-312-841-1	Sequence 1, Appli
13		50.4	4.1	13427	15	US-10-311-455-1900	Sequence 1900, Ap
14	c	50	4.1	8951	15	US-10-311-455-768	Sequence 768, App
15		50	4.1	16439	15	US-10-311-455-860	Sequence 860, App
16		49.8	4.1	494	16	US-10-424-599-18680	Sequence 18680, A
17	c	49.8	4.1	5442	15	US-10-311-455-1942	Sequence 1942, Ap
18	c	49.8	4.1	10328	15	US-10-311-455-1517	Sequence 1517, Ap
19		49.4	4.1	2000	9	US-09-938-842A-2755	Sequence 2755, Ap
20	c	49.4	4.1	2000	9	US-09-938-842A-2909	Sequence 2909, Ap
21		49.4	4.1	2000	11	US-09-938-842A-2755	Sequence 2755, Ap
22	c	49.4	4.1	2000	11	US-09-938-842A-2909	Sequence 2909, Ap
23	c	49.4	4.1	7676	15	US-10-240-485-151	Sequence 151, App
24		49	4.0	8056	18	US-10-473-126-240	Sequence 240, App
25	c	48.6	4.0	6129	15	US-10-240-485-135	Sequence 135, App
26		48	3.9	1501	18	US-10-473-126-182	Sequence 182, App
27		48	3.9	3673778	15	US-10-312-841-1	Sequence 1, Appli
28	c	47.8	3.9	16724	15	US-10-311-455-1064	Sequence 1064, Ap
29	c	47.8	3.9	16724	15	US-10-240-485-90	Sequence 90, Appl
30	c	47.6	3.9	16724	15	US-10-311-455-1063	Sequence 1063, Ap
31	c	47.6	3.9	16724	15	US-10-240-485-89	Sequence 89, Appl
32	c	47.4	3.9	17421	14	US-10-239-676-54	Sequence 54, Appl
33	c	47.4	3.9	17421	15	US-10-240-453-56	Sequence 56, Appl
34	c	47.4	3.9	19634	18	US-10-473-126-302	Sequence 302, App
35	c	47.2	3.9	397	14	US-10-198-846-1348	Sequence 1348, Ap
36	c	47.2	3.9	6749	16	US-10-221-714A-247	Sequence 247, App
37	c	47.2	3.9	17703	16	US-10-257-166-34	Sequence 34, Appl
38	c	47	3.9	6816	18	US-10-723-860-7676	Sequence 7676, Ap
39	c	47	3.9	18988	15	US-10-311-455-674	Sequence 674, App
40	c	47	3.9	18988	15	US-10-240-485-62	Sequence 62, Appl
41	c	47	3.9	18988	16	US-10-221-714A-64	Sequence 64, Appl
42	c	46.8	3.8	83391	17	US-10-433-793-124	Sequence 124, App
43		46.8	3.8	102980	13	US-10-087-192-1330	Sequence 1330, Ap
44	c	46.6	3.8	5163	15	US-10-311-455-1222	Sequence 1222, Ap
45	c	46.6	3.8	15373	15	US-10-311-455-440	Sequence 440, App

ALIGNMENTS

RESULT 1
US-10-473-126-386
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
; LENGTH: 8056
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

Query Match 4.8%; Score 59; DB 18; Length 8056;
Best Local Similarity 44.3%; Pred. No. 0.06;
Matches 483; Conservative 0; Mismatches 585; Indels 23; Gaps 5;

```
Qy        101 ATTTATTTACAAGGATACTGAATATGACTGCACTGAATATGCTAAATCAAATAAGCATCC 160
              |||| || | ||  | |   || || | |    |   | |   || | |  |     |
Db       2396 ATTTTTTGAAAATTAAAAAAAAAAATAATTTATTTATTTTTAAAAATTAATTTTTTTTTAT 2455

Qy        161 TGGCGGTCTTAATTTCCTCAATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTT 220
              |      |  ||  ||  | | | | ||  || |  | || ||| || | | |||| | |
Db       2456 TTTTATTTATATATTTTTTATTATTTTGTTTTAAAATAATTAAAATATTAATGAAAAATA 2515

Qy        221 CAGA-----ACACTCCATTCTAAGTAGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACT 275
               | |     | | |  ||| |||      ||  ||||||| |||||        ||   |
Db       2516 AAAAAAAATAAATTAAATTATAAAATAAATTAAAAAATTTAAAAATTTAAATAAAATTTT 2575
```

```
Qy       276 GGCGCAAAATAAGAGGAGACTGAATCTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTT 335
Db      2576 TAAAAAATAAAAAATAAATGAAAAAAAAAATTATAAATATAAATAATTTAAAAAAAAAAA 2635

Qy       336 AAGCATTTATTCGAACCAAACTGGCCTATCGAAATTGGTTTATTCTTAACTACCTTTACT 395
Db      2636 ATAAAAAAAAATGTTAAAATTTAAAATTTTAAAAATAATAAATTTTTAAATTTTTTTAAA 2695

Qy       396 TTATTTGTCACTGGATGTTTGACTCAAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTA 455
Db      2696 TTTATTATTTTTTTTTTTTAATAAAAATAATTTGTTAAATAAAT----------TATTA 2744

Qy       456 ATGCAAATCATCAGTGGTTGGATTGGTCACTCTATGAACCACAATCGTAACCCTATATTA 515
Db      2745 ATAAAAATAAAAAAAAATATTTAATTGTTTATTTTTAATAAATTTAAATGAATTTTGTTTA 2804

Qy       516 AGAAAATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGT 575
Db      2805 TTAAAATTGTTATTAATAATAAATGATATTTTTTTATAAAATATTAAAAATTTGAAATTA 2864

Qy       576 AGGAAGCACAATCAACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAA 635
Db      2865 TTTTATGAAAAAAATTTTGTTGAAAAAAATAAAAAATTTGAAAATGAAAAAAAAATGTTA 2924

Qy       636 CACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATC 695
Db      2925 AAATTTTATGAAATGTAAAATTTAAATGAATTATTATTGTTAAAAAAATAAAAAAAAATT 2984

Qy       696 TTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTT---GCACTGGGTATTATTA 752
Db      2985 ATTGAAAAATTGATATATTTTAAAATGATGATGTTTAAATTTTATAAATATAAATTAAAA 3044

Qy       753 TTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCT 812
Db      3045 ATAAAATAAATGTTATAATTGTGAATGAAATTTGAATT-ATTGTTATTTTTATGTGTAAT 3103

Qy       813 TCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCA 872
Db      3104 TTAATAAAAAATTAATTTTTTTAAAATGAATTTAATGATGATTGAAAAATTGTTTTATAT 3163

Qy       873 TTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTA 932
Db      3164 ATAATTTTAAATAATAATATTTTTAATAAAAATTTTGTTGT---TTGAATTGTAAATAAA 3220

Qy       933 CTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTC 992
Db      3221 ATTTTAATTTATTATATTTTTTTTTAATAATTTGATTATTATTTTTAATTAATTGATTGT 3280

Qy       993 TACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAG 1052
Db      3281 AAAAAAATTTTTGTTATTGTTGTTGTTATTTTTTTGAAAAAAAAAAAAATATAATTGTAAT 3340

Qy      1053 ATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAA 1112
Db      3341 GTAAAAAAATGAATAAAATTTGAAATGATGATTAAAAAATAAAAATGAAATAAATAATAA 3400

Qy      1113 ATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTA 1172
Db      3401 TTTGAATTTTTTTTAAATATTTTTGTTAAATGTTGAAAATTATAATTTGTTTTATTATTA 3460

Qy      1173 TTACTTTTTAA 1183
Db      3461 TTATTAATAAA 3471
```

RESULT 2
US-10-311-455-1663/c
; Sequence 1663, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian

; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1663
; LENGTH: 6103
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1663

Query Match 4.8%; Score 58; DB 15; Length 6103;
Best Local Similarity 46.6%; Pred. No. 0.086;
Matches 221; Conservative 0; Mismatches 250; Indels 3; Gaps 1;

```
Qy       748 TATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCA 807
Db      1844 TCTACTTAAAACAAAATCTAACTACCTTTAAAACATCTCTACAAAATCCTTTACTCCTAA 1785

Qy       808 GTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAAT---GAAATTCGAAAGAAGAATCA 864
Db      1784 AAACTACTAATTTTATCCAAAACAAAATTAACCAACTATAAAAATCAAAACCAAAAAAAA 1725

Qy       865 CTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATAT 924
Db      1724 ACAACTCCTTTTCAAATCACACAACAAATTAAAAACTAAACAAAAATTATAACTTAAATC 1665

Qy       925 TCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAA 984
Db      1664 TTTTACTACATCGTATAATAACTTATTACATATCATCTCCACCACACCCTCTTTAATAAA 1605

Qy       985 TTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGA 1044
Db      1604 TCCCATAAAACAAAATACTCCAATCTCCAACTCATTAAAATACATAAAAAAAAAACAATA 1545

Qy      1045 ACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATT 1104
Db      1544 AAAAAAATAAACACTTTTTAACAAATAATTATAAAAAAAAAACCTTTCCATACAAACATAT 1485

Qy      1105 TATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTG 1164
Db      1484 AATAATAAATACATCTATTATAAAAAATCCTACATAAAAAAAAAAATTAAATATACTAAA 1425

Qy      1165 TTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1218
Db      1424 TACGTTCATTTTTCTAATCAAAAATTAAAACATATTAAAAAATAAAAACAAAAA 1371
```

RESULT 3
US-10-473-126-240/c
; Sequence 240, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of hematopoietic cell
; TITLE OF INVENTION: proliferative disorders
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/473,126
; CURRENT FILING DATE: 2003-09-26
; NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 240
; LENGTH: 8056
; TYPE: DNA

```
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-240

  Query Match             4.7%;  Score 57.4;  DB 18;  Length 8056;
  Best Local Similarity   44.3%;  Pred. No. 0.13;
  Matches  283;  Conservative    0;  Mismatches  351;  Indels    5;  Gaps    1;

Qy        586 ATCAACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACA 645
              || || |   ||| |    ||  |   || | ||  | || |   |||    | | |
Db       2746 ATTAATAATTTATTTAACGAATTATTTTTATTAAAAAAAAAAAAATAATAAATTTAAAAAA 2687

Qy        646 AATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTT 705
              | ||       |    || | |  ||| || |   ||      |  | ||   || ||
Db       2686 ATTTAAAAATTTATTATTTTTAAAATTTTAAATTTTAACGTTTTTTTTTATTTTTTTTTT 2627

Qy        706 ATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACT 765
              | |  ||||  |   ||  | ||   ||     |  | ||||| || ||  |
Db       2626 TTAAATTATTTATATTTATAATTTTTTTTTTCGTTTATTTTTTATTTTTTAAAAATTTTA 2567

Qy        766 TCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTG 825
              | || ||  | |   ||| | | ||  ||  |  |||  |   |  || | |  || | |
Db       2566 TTTAAATTTTTAAATTTTTTAATTTATTTTATAATTTAATTTATTTTTTTTTATTTTTCG 2507

Qy        826 GAAA-----TCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGA 880
              ||     | || |   | ||| |||||   ||| || |   |  | |  ||       |
Db       2506 TTAATATTTTAATTATTTTAAAACGAAATAATAAAAAATATATAAATAAAAATAAAAAAA 2447

Qy        881 ACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTAT 940
              | | |      |   | ||| ||   |||     | |         || | |   |
Db       2446 AATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAATTTTCGAAAAATAATAAAATT 2387

Qy        941 GGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATT 1000
              |    | |    | |   || || |  |||    || || |    | ||   |
Db       2386 TAATAAATTATTTATAAAAATAAAAAATTATATTTAAAAATTAAATAAAATTTATTTAAA 2327

Qy       1001 ACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGA 1060
                | |||  | |    || ||   || || ||||  |    |    |   ||
Db       2326 TACGAAAATTAATAAATTATTTTAAAAATAAATAAATTAAAAAATTAAAATTTTTTTAAT 2267

Qy       1061 AGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATAT 1120
                 ||  | | |||||    |  |||   || |  || ||| ||||   |  || |
Db       2266 TTATTTTAAATATTAAAAAATAAATTTTTATTAAAATTAAATTTTTTTTTTTTTTTATTT 2207

Qy       1121 TTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTT 1180
              |||  ||      |  | | | || ||  |||  | | ||  |||  |||  |||||  || |
Db       2206 TTTTTTTCGTTTTTTTAAATTTTTAATAAAATTATAATTTTTTTTTTTTTTTATTTTTAT 2147

Qy       1181 TAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1219
              |  |    | || | |   |||  ||||  | |  ||| |
Db       2146 TTTTAATAAAATAAAAAATTATTAAAATTATATTAAATA 2108


RESULT 4
US-10-473-126-386/c
; Sequence 386, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 386
;   LENGTH: 8056
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-386

  Query Match             4.5%;  Score 54.6;  DB 18;  Length 8056;
  Best Local Similarity   45.3%;  Pred. No. 0.5;
  Matches  282;  Conservative    0;  Mismatches  334;  Indels    7;  Gaps    2;

Qy        596 TATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTA 655
              ||| |  | | |  | ||  || | |    |   |||    | ||  ||||     ||
Db       6415 TATATATATATAATATATAATATATAATTTTTTTTAATAAATATAATAAAAATTTATTTA 6356

Qy        656 ATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATT 715
                | | || ||    |  |  |  |||         |  |||  ||  |  | | || |
Db       6355 TAACACATTTTTTAAATATACATAATAATACATTATTAAATTATTTTAATTATTTTAAAT 6296

Qy        716 TGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGT 775
                |   | ||  |    ||  | |   | || ||||| |  | ||  || || ||  |
Db       6295 ATCATTTTTTTTTTTATATTTTTATTATTTTTTTTTTATTTATATTAATATTTTACATTTT 6236

Qy        776 AATTCTTTCTGAATT--GATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCAT 833
              ||        |  |||   |||  |  ||| || | |  || ||    ||  |   |
Db       6235 AAACACACATTTATTAATATTTTTTTTTTTTTTATTTTTTTTAAAATTTTTAATTAAAAC 6176

Qy        834 GAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCT 893
                || |    |   | |  || || |    | ||   | | |  |||| | ||| |
Db       6175 ACAATTTTTTTTTTAATTTTAACAAAACATTTTATAATTTAATACACACATAACATAACA 6116

Qy        894 GCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAA 953
                || || | | |||  |||| |   | |     |   | |||     |  | | ||
Db       6115 ATAAACACATAATACATTTTCAATATAAAAAATAATTATTTCATTTACAATAATTTATAT 6056

Qy        954 TATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGT 1013
              |  || | | ||  ||     |||    |  |||     ||  | | |     |||
Db       6055 TTATAAAATAAAAATACAATTTATTAAATAATATTATTACAAAAAAATAATAAAAAAAAA 5996

Qy       1014 GTCATAATTGCTGAAGAATTAAAGA-----AGTGGAACCTTAAGATTCATGAAGGACCTA 1068
                | ||||     || ||  |||| |     | | |  | |  | ||| ||  |    |
Db       5995 AAAAAAATTTAAAAAAAAATAAAAAAAAATAATAAACACACATAATTAATTTATACAATT 5936

Qy       1069 TTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAG 1128
              |||||      || |    ||| ||| ||||  | |  ||| | |   | ||||| ||
Db       5935 TTTTTATTTTATTTATATTTTAAAATTAAAATATTCAATTATTCAATTAAATTTTTTTTT 5876

Qy       1129 TAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGA 1188
              || |   |  | || |   |  | | ||||      |||  | |   |||  |||   |
Db       5875 TATTTTTTTTATTTTTTTTATTTTTTTTATAAACATTTTTTTCAAATTTTTAAAATAAAA 5816

Qy       1189 GAAAACAGTCTTAACAAAAAAAA 1211
                | |  |      || |||||||||
Db       5815 AACATAAAAAAAAAAAAAAAAAA 5793


RESULT 5
US-10-473-126-328
; Sequence 328, Application US/10473126
; Publication No. US20040234973A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Methods and nucleic acids for the analysis of  hematopoietic cell
;  TITLE OF INVENTION:  proliferative disorders
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/10/473,126
;  CURRENT FILING DATE:  2003-09-26
;  NUMBER OF SEQ ID NOS: 1258
; SEQ ID NO 328
;   LENGTH: 1501
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-473-126-328

  Query Match             4.4%;  Score 53.2;  DB 18;  Length 1501;
  Best Local Similarity   46.1%;  Pred. No. 0.49;
  Matches  259;  Conservative    0;  Mismatches  293;  Indels   10;  Gaps    2;
```

```
Qy        663 TTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGA 722
Db        900 TTAATGATTTTTAATTTTAAATTTAAATTAATTTTGATGTTTGTTTAATTATTTGTTGAT 959

Qy        723 ATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTT 782
Db        960 TTGTTTTTTAAATTTAATAATTTAAATTTAAATAATATTGAAATGAAAATTTTGAAATTT 1019

Qy        783 TCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAA 842
Db       1020 AAAATTTGAAATGTAATTTTTTTAAATTTTAATTTTATT----TATAAAATAAAAAAATA 1075

Qy        843 ATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGA 902
Db       1076 ATAATATTTATTTTATAAAATTATTATAAAATTTTAAAAAAATAAAAATATTTAAAAAAA 1135

Qy        903 AACTACGCTTTCCACGACATATTCTCTCTACTT------ATTATGGGTGGTATGTAATAT 956
Db       1136 AAATTTAATTTAAATAAAAAAATAAATGATGTTAATTTAATTTTTGTTTTAAAAAAATGA 1195

Qy        957 TAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTC 1016
Db       1196 TATAAAAATAAAAAATTTTTTTTAAAATAAAATTAATTAAAATTTGTTATATTTTTTTTT 1255

Qy       1017 ATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAA 1076
Db       1256 ATTTTATATAAAATATTATAAAAATTATATAATAAAAAATAAAAATTTAATATTTAAAAA 1315

Qy       1077 AAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTA 1136
Db       1316 ATATTATAAAATTAAATTATTTTAAATTTTATTTTATATTATTAATTAAAATAATAAATA 1375

Qy       1137 ACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTAATCTGAGAAAACAG 1196
Db       1376 ATTTTTTTATATAATTTTTTTTTTTAAATTTTAAATATATTTTATTAAAAAATAAAATTT 1435

Qy       1197 TCTTAACAAAAAAAAAAAAAAA 1218
Db       1436 ATTTAATGTTAAAATTAAAATA 1457
```

RESULT 6
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 4.3%; Score 52; DB 9; Length 640681;
Best Local Similarity 54.8%; Pred. No. 11;
Matches 103; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy       1031 ATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTG 1090
Db      95585 AGTAAAAAAATATTAATATACTATCAATAATTGAAATATTTTTTAAAAATCTTTTTCTAT 95644

Qy       1091 AAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAA 1150
Db      95645 TTAAAAAACATATTTTTTATATATCTGACTAGTATTTTTTAAAAAAATCATTTTAATATT 95704

Qy       1151 TGTGTTATGGTTTGTTTACTTATTACTTTTAATCTGAGAAAACAGTCTTAACAAAAAAA 1210
Db      95705 TTTTGATATTAAAATGATTTTTTATTCTTAATATATATTATTCAGTTATATAAAATAA 95764

Qy       1211 AAAAAAAA 1218
Db      95765 AAAATATA 95772
```

RESULT 7
US-09-960-352-11218/c
; Sequence 11218, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11218
;   LENGTH: 424
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 48-LIB3058-026-Q1-K1-D12
US-09-960-352-11218

Query Match 4.2%; Score 51.8; DB 9; Length 424;
Best Local Similarity 63.0%; Pred. No. 0.56;
Matches 80; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy       1093 AATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATG 1152
Db        194 AAAATTTAATTTATTTTATTTTAAATTTTTTTATTTTTTTATTTTTTTTTTTTTTT 135

Qy       1153 TGTTATGGTTTGTTTACTTATTACTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAA 1212
Db        134 TATTTTTTTTTTTTTATTTTTTATTTTTAATTTAAAAAAAAAAAAAAAAAAAAAAAAAA 75

Qy       1213 AAAAAAA 1219
Db         74 AAAAAAA 68
```

RESULT 8
US-10-311-455-2017/c
; Sequence 2017, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2017

```
;   LENGTH: 7319
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2017

  Query Match             4.2%;  Score 51.8;  DB 15;  Length 7319;
  Best Local Similarity   48.7%;  Pred. No. 1.9;
  Matches  170;  Conservative    0;  Mismatches  177;  Indels    2;  Gaps    1;

Qy          3 TAAGCAAACTAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAATT 62
              ||| | |  |||||  | | || || ||||      || | || ||  ||  | ||| |
Db       1875 TAAACCACTTAAATCCATATAAAAATCATTTAAAATTTTAACAAAACACAAACAAAAAAT 1816

Qy         63 GTTCTTGAAAATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTGAA 122
              |||| |||||| || | | ||  | | |    | ||    || || ||  ||||  ||
Db       1815 TATCTTTAAAATATACACAACCTAATAATCATTAAAAACAAAATCTAAAACTATACACAA 1756

Qy        123 TATGACTGCACTGAATATGCTAAATCAAA--TAAGCATCCTGGCGGTCTTAATTTCCTCA 180
              ||    | |  | |    |||  || |  ||| ||||      |   ||||||  |
Db       1755 TAATTACCCCCAAATTTAATAAAACCATAATTAACCATCAATTATTTAAAAATTTCACCG 1696

Qy        181 ATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGT 240
              |  || |||  | | ||  |    |  | | | | | | ||| || |  || |
Db       1695 CATAATTCATTTCTAAAAAAAATATATAAATTATAAAATCAATATCACAACAATAAAATT 1636

Qy        241 AGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAAT 300
              ||||    |  |  ||||  | |||    ||||       | |   | |
Db       1635 TAAAAAATAAAAAAAAAATCCTTAATCCCACAAACTAATACAAATATCTTTTAAATATATT 1576

Qy        301 CTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGA 349
              |  | | | | | | |  | | |||| || ||    ||    | ||||  |
Db       1575 CACTATATATAATTTTTACAATAAAAAACAAAAAAAACTTATAATTCTA 1527


RESULT 9
US-10-311-455-74/c
; Sequence 74, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 74
;   LENGTH: 5728
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-74

  Query Match             4.2%;  Score 51.4;  DB 15;  Length 5728;
  Best Local Similarity   54.7%;  Pred. No. 2;
  Matches  123;  Conservative    0;  Mismatches  101;  Indels    1;  Gaps    1;

Qy        989 TTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCT 1048
              | |  ||  |  | | |||   | | ||| ||      || ||  ||| ||       |
Db       2910 TCTAAACGAAAAAACGAAACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAAATTCTTT 2851

Qy       1049 TAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATT 1108
```

```
              |||  ||||| ||     | ||      ||| |||  | || || ||||| ||||  |||
Db       2850 TAAATTTCATCAATTTTATTTTAAAACTAAAACTCTTCATTAAATATAAAAAAATCAATT 2791

Qy       1109 TTAAATGCATATTTTATTAGTA-ATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTT 1167
                ||||   ||||  || |  | | | | | ||||  |   ||||  | ||    |||
Db       2790 CAAAATTTCTATTCAATCATAATAAATTTAAAATTTTTTTAAATTTCTTTATATTTAACA 2731

Qy       1168 ACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAA 1212
              |  |||||| || |||| | | || | |  || || |||  || |
Db       2730 ATATATTACATTATAATATTATAATAAAAACTAAATAAACTAATA 2686


RESULT 10
US-10-311-455-1982/c
; Sequence 1982, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1982
;   LENGTH: 5750
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1982

  Query Match             4.2%;  Score 51.2;  DB 15;  Length 5750;
  Best Local Similarity   45.7%;  Pred. No. 2.3;
  Matches  179;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
              ||||  | |||||  |||  ||   |  | |        |         || |||| ||||  |
Db       4820 AAATATTTAAAATAAAAATAACGACACTAAATCTTAAAAAAATAAATCTTTCAAACAAAA 4761

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGG 946
              || |  |  || ||  || ||  |      |  | ||   |||| | |
Db       4760 AACAAATATAAACAATAAATATCCCACGAAAAAAATTCCCTCTAAAATACGCCCTAAAAA 4701

Qy        947 TATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAA 1006
              | | || | ||   |  | |  |||| ||   | || || || | |  | ||  | |
Db       4700 AACGAAAAAATATTTTATAAAAAACTTATTTTAAAAACTTTCTCTTTCTAAATATCTCCT 4641

Qy       1007 AGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACC 1066
              || | |   || |  | ||   || |      |  |||| || ||   | |
Db       4640 TTATCCTATTTAAAATTATTAAACCTTCATACTATTTAACCTCAAAATATTTAACTACTT 4581

Qy       1067 TATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATT 1126
              |  | |  |||||   ||||  | |  |   | ||    ||   |  |   | ||   ||
Db       4580 AAAATATTCAAAATTAAACCTATAATCAATCAAAATAAAATAAAATCTATTTCTTAATTT 4521

Qy       1127 AGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCT 1186
              | | ||       | ||  |      ||  |  | |  ||  ||   ||  |||   |  |||||
Db       4520 ATTTATTTATTTATTTTTTTTTTATTTTATTTATTTATTTATTTTTTTAAAATAAAATCT 4461

Qy       1187 GAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1218
                    |  | |  |||  |||| | || || |
Db       4460 CGCCCTATCGCCCAAACTAAAATACAATAACA 4429
```

```
RESULT 11
US-10-221-714A-434/c
; Sequence 434, Application US/10221714A
; Publication No: US20040048254A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with
;  TITLE OF INVENTION:  tumor suppressor genes and oncogenes
;  FILE REFERENCE: 5013.1005
;  CURRENT APPLICATION NUMBER: US/10/221,714A
;  CURRENT FILING DATE:  2003-01-21
;  PRIOR APPLICATION NUMBER: PCT/EP01/02955
;  PRIOR FILING DATE: 2001-03-15
;  PRIOR APPLICATION NUMBER: DE 10013847.0
;  PRIOR FILING DATE: 2000-03-15
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 540
; SEQ ID NO 434
;   LENGTH: 5750
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-434

  Query Match             4.2%;  Score 51.2;  DB 16;  Length 5750;
  Best Local Similarity   45.7%;  Pred. No. 2.3;
  Matches  179;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
              ||||  | |||||  |||  ||   |  | |       |        || |||| ||||  |
Db       4820 AAATATTTAAAATAAAAATAACGACACTAAATCTTAAAAAAATAAATCTTTCAAACAAAA 4761

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGG 946
              ||  |   ||  ||  ||  ||  |      |  | ||   ||||  | |
Db       4760 AACAAATATAAACAATAAATATCCCACGAAAAAAATTCCCTCTAAAATACGCCCTAAAAA 4701

Qy        947 TATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAA 1006
              | | ||  |  ||    |  | |  |||| ||   | ||  ||  ||  | |  |  ||  | |
Db       4700 AACGAAAAAATATTTTATAAAAAACTTATTTTAAAAACTTTCTCTTTCTAAATATCTCCT 4641

Qy       1007 AGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACC 1066
                ||  | |   ||  |  | ||   || |      |  |||||  ||  ||   | |
Db       4640 TTATCCTATTTAAAATTATTAAACCTTCATACTATTTAACCTCAAAATATTTAACTACTT 4581

Qy       1067 TATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATT 1126
              | | |  |||||   ||||  | |  |   | ||    ||   |  |   | ||   ||
Db       4580 AAAATATTCAAAATTAAACCTATAATCAATCAAAATAAAATAAAATCTATTTCTTAATTT 4521

Qy       1127 AGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCT 1186
              | | ||      | || |      || |  | |  ||  ||   ||  |||   |  |||||
Db       4520 ATTTATTTATTTATTTTTTTTTTATTTTATTTATTTATTTATTTTTTTAAAATAAAATCT 4461

Qy       1187 GAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1218
                 |  | |  |||  ||||  | ||  ||  |
Db       4460 CGCCCTATCGCCCAAACTAAAATACAATAACA 4429


RESULT 12
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
;   LENGTH: 3673778
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (3294164)
US-10-312-841-1

  Query Match             4.2%;  Score 51.2;  DB 15;  Length 3673778;
  Best Local Similarity   47.0%;  Pred. No. 34;
  Matches  190;  Conservative    0;  Mismatches  213;  Indels    1;  Gaps    1;

Qy        815 TATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATT 874
              | ||| | ||  ||  | |  ||||  ||   | | ||   |  | |  || |||   |||
Db    1656921 TTTTCATATTCAAATTTAAAAAAAAAAACCTATAATCACTAACATATCCAATTTCTAATT 1656862

Qy        875 TTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACT 934
              ||||  ||  |  |  |      | |  | ||   ||  |||  |  | |   | | |  | |
Db    1656861 TTTCAAATAAAATTTCTTATCTACAACAATAAACATTTTAAATAAAACTATTTTTAAAAT 1656802

Qy        935 TATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTA 994
              ||       |  | | |||    |  || |   ||  |  | |  || |||   | || |
Db    1656801 TAAATACATTTATTTTTAAATAAAACCTAATTTTCTATAATCCATATCAATAATTCTCAA 1656742

Qy        995 CAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGAT 1054
                || ||  ||   |    | |||   |   | |||  |||  |  |        || ||
Db    1656741 ACTCTTTCCATAAAAAC-ACACTTAAAAACGACAAAATAAAACATTTTTCTTAAAAACAT 1656683

Qy       1055 TCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAAT 1114
                |   |   |  |  | ||||    | ||    ||  ||||.|  || | ||| | |
Db    1656682 AAACTTAATCTCATTAACTCAAAACAAACTCCAAATAACTTAAAAATTTTAAATTTTATT 1656623

Qy       1115 GCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATT 1174
                |||| ||| |  ||||  || |  |  ||  ||   | |        | || |  || |
Db    1656622 TTATATATTAATTTTAATTTTACCTTTAATATCAATCATCTATATAAATTTTAAAATAAT 1656563

Qy       1175 ACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1218
                || ||||   | ||||  |     |  | ||  ||| |||||
Db    1656562 TTCTTCTAATTCTATAAAAAATATCAATAATAATTAAATAAAAA 1656519


RESULT 13
US-10-311-455-1900
; Sequence 1900, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determi
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1900
;   LENGTH: 13427
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
```

```
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1900

  Query Match             4.1%;  Score 50.4;  DB 15;  Length 13427;
  Best Local Similarity   45.3%;  Pred. No. 4.7;
  Matches  223;  Conservative    0;  Mismatches  266;  Indels    3;  Gaps    1;

Qy        707 TTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTT 766
Db       3718 TTGTTAATTTTAATGTAGATAATTAGTAGGTTGATTTTTATTTTTTATTTTATTTTATTT 3777

Qy        767 CTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGG 826
Db       3778 TTAAGTATTATATATTTTATTTTTTAAGTATTATAGTAGAAGGAGTTTATTTTTTAGTTG 3837

Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
Db       3838 AAATTATGTATGTTGATGTGAGTATAAATAATTTGTGGTATTAGAGATTTTAATTAAAAT 3897

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTT---CCACGACATATTCTCTCTACTTATTATGGG 943
Db       3898 ATTTTTTTTAAAGAAATATTATAATTTTTAATAAGTTATATTATTTTTATATGATATATA 3957

Qy        944 TGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACC 1003
Db       3958 TGTTTTGTAAGTTTATATAAAATTTAAATTGGTTAAATAGAAAATGTTGTAAAATTTTTG 4017

Qy       1004 CAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGG 1063
Db       4018 AAAAGTTATTTGTTTAATTTTATTGAATTTTAATTTTGGATTGTAAAAATAAGATAATG 4077

Qy       1064 ACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTT 1123
Db       4078 ATAGTTAATTTATAGGGTTTATTATGGATGAAATTAAATATTGTATGTAAAAGTATTTT 4137

Qy       1124 ATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAA 1183
Db       4138 GTAAATGAAAAAAAAATTTTTTTTAAATGTAGATTATTAATATTTTAAATTTTTTATAAA 4197

Qy       1184 TCTGAGAAAACA 1195
Db       4198 GTTTAATATAGA 4209

RESULT 14
US-10-311-455-768/c
; Sequence 768, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 768
;   LENGTH: 8951
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-768

  Query Match             4.1%;  Score 50;  DB 15;  Length 8951;
  Best Local Similarity   44.1%;  Pred. No. 4.8;
  Matches  209;  Conservative    0;  Mismatches  265;  Indels    0;  Gaps    0;

Qy        746 ATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTT 805
Db       1019 ATTACAAAAAAAATACAAAATAACAATCCTCCCTCTAACTTAATAAAAATATATTTTTC 960

Qy        806 CAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCAC 865
Db        959 CATAAAACATATATATATTTTACAAATAAAAAATAATTTAAAATAAAAACATATATTTC 900

Qy        866 TTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATT 925
Db        899 TACTCTTTAAATTAAAAAAAACTTAAAATCTTTCAAATTCAATACTTCCCAAAAATAATT 840

Qy        926 CTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAAT 985
Db        839 TTCCTTTTAATTAAAATTTTATTCTTATATTACTTTTTACTATTATTTTTACAATTTACC 780

Qy        986 TCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAA 1045
Db        779 TAATACTAATAAAATCTCAAAAACTATATTTAATATTAAAATATAATTTTTCCAAAAAAT 720

Qy       1046 CCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTT 1105
Db        719 AACAAATAATTAATAATCTCCCCTTTTCCTCAACAACATAATTATACCCAAAAAAAATTT 660

Qy       1106 ATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGT 1165
Db        659 ATCTTTAAAAAAATATAATTTATAAAAAAAAAAAAAAATAAATAAATAAATAAAATATAA 600

Qy       1166 TTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1219
Db        599 TACAATTTTAAAAATAAATTTTTAAATAACCCAAAAAATAAAAAACCTACACAA 546

RESULT 15
US-10-311-455-860
; Sequence 860, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
;  TITLE OF INVENTION:  cytosine methylation
;  FILE REFERENCE: 5013.1014
;  CURRENT APPLICATION NUMBER: US/10/311,455
;  CURRENT FILING DATE:  2002-12-16
;  PRIOR APPLICATION NUMBER: PCT/EP01/07537
;  PRIOR FILING DATE: 2001-07-02
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 860
;   LENGTH: 16439
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-860

  Query Match             4.1%;  Score 50;  DB 15;  Length 16439;
  Best Local Similarity   50.0%;  Pred. No. 6.3;
  Matches  155;  Conservative    0;  Mismatches  150;  Indels    5;  Gaps    1;

Qy        912 TTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCAC 971
Db       5272 TTATAAAAAATTTTTTAGGTAGTTTTTTTTTGTTTTAGGAATATTTATGTTAATTTATT 5331
```

```
Qy        972 TTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGT-----GTCATAATTGCTG 1026
              ||||    | | |||     | |  |||| |    | | | ||       | | ||||  |
Db       5332 ATTTTTAGAGATATTAGAATATGAAGATAAGAGTATGATAGTTTTCGTTTAAAATTTTTT 5391

Qy       1027 AAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACC 1086
              ||    | |  | |  | ||  |      |        |||  | |  ||  ||    | |
Db       5392 AATGGTAATTGGAAGGAAAGGTATTTTTGGGGTTTGGATTTTTAATTTTAATTGGTTAAT 5451

Qy       1087 TTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAG 1146
              ||| ||||| | ||| ||| |||||     ||||||| ||   || | | | || | |
Db       5452 TTTTAAAATTATTAAGTTTTTTTTATTATTATATTTTTTTTTAAAGATTGAAAAATATTT 5511

Qy       1147 GAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAA 1206
                |||   || |   || ||| |  |    |||  ||  |  | | |       || |||
Db       5512 ATAATTATTTTTTTGTTATTTTCGGAAATGTTTGGAAGGTAGGTAGAGTTAAAAAAAAAA 5571

Qy       1207 AAAAAAAAAA 1216
              ||||||||||
Db       5572 AAAAAAAAAA 5581
```

Search completed: January 19, 2005, 17:17:33
Job time : 772 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 19, 2005, 12:54:40 ; Search time 371 Seconds
(without alignments)
2335.449 Million cell updates/sec

Title: US-10-070-666A-1
Perfect score: 1219
Sequence: 1 agtaagcaaactaaatttaa.....taacaaaaaaaaaaaaaaaa 1219

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		58.4	4.8	1141	4	US-09-806-708B-22	Sequence 22, Appl
2	c	54.4	4.5	1141	4	US-09-806-708B-22	Sequence 22, Appl
3		52	4.3	640681	4	US-09-790-988-1	Sequence 1, Appli
4		47.8	3.9	832	4	US-09-621-976-2813	Sequence 2813, Ap
5	c	45.6	3.7	640681	4	US-09-790-988-1	Sequence 1, Appli
6		45.4	3.7	1436	4	US-09-614-912-35	Sequence 35, Appl
7		44	3.6	11049	4	US-10-204-708-22	Sequence 22, Appl
8		43.8	3.6	2949	3	US-09-412-554A-3	Sequence 3, Appli
9	c	43.8	3.6	51952	3	US-08-947-823-1	Sequence 1, Appli
10	c	43.4	3.6	826	4	US-09-270-767-29614	Sequence 29614, A
11	c	43.4	3.6	1348	4	US-09-270-767-13608	Sequence 13608, A
12	c	43.4	3.6	2999	4	US-09-710-279-3580	Sequence 3580, Ap
13	c	43.4	3.6	3099	4	US-09-710-279-4201	Sequence 4201, Ap
14		43.4	3.6	3501	4	US-09-710-279-3614	Sequence 3614, Ap
15	c	43.4	3.6	4015	4	US-09-710-279-3619	Sequence 3619, Ap
16		43.4	3.6	4207	4	US-09-710-279-3631	Sequence 3631, Ap
17		43.2	3.5	2141	4	US-09-270-767-25196	Sequence 25196, A
18		43.2	3.5	3016	4	US-09-270-767-9898	Sequence 9898, Ap
19	c	43	3.5	1422	1	US-08-319-704-5	Sequence 5, Appli
20		42.4	3.5	10640	4	US-09-417-485D-5	Sequence 5, Appli
21	c	41.4	3.4	665	3	US-08-998-416-937	Sequence 937, App
22		41.4	3.4	3618	1	US-07-872-678A-36	Sequence 36, Appl
23		41.2	3.4	6801	4	US-10-204-708-62	Sequence 62, Appl
24		41	3.4	5852	1	US-07-867-106-2	Sequence 2, Appli
25		40.8	3.3	526	4	US-08-956-171E-516	Sequence 516, App
26		40.8	3.3	526	4	US-08-781-986A-516	Sequence 516, App
27		40.8	3.3	1055	4	US-09-806-708B-23	Sequence 23, Appl
28		40.8	3.3	1905	4	US-09-270-767-10264	Sequence 10264, A
29	c	40.8	3.3	20674	4	US-09-641-638-651	Sequence 651, App
30	c	40.8	3.3	20674	4	US-10-170-097-651	Sequence 651, App
31	c	40.6	3.3	58407	4	US-08-916-421B-2	Sequence 2, Appli
32	c	40.6	3.3	58407	4	US-09-692-570-2	Sequence 2, Appli
33	c	40.4	3.3	19124	2	US-08-487-826B-13	Sequence 13, Appl
34	c	40.2	3.3	837	3	US-08-998-416-288	Sequence 288, App
35	c	40	3.3	665	2	US-08-883-795A-36	Sequence 36, Appl
36	c	40	3.3	832	4	US-09-621-976-2813	Sequence 2813, Ap
37	c	40	3.3	1055	4	US-09-806-708B-23	Sequence 23, Appl
38		40	3.3	2058	2	US-08-749-391-1	Sequence 1, Appli
39		40	3.3	2058	3	US-09-390-200-1	Sequence 1, Appli
40	c	40	3.3	162450	3	US-09-345-882-1	Sequence 1, Appli
41	c	39.8	3.3	319608	4	US-09-539-333D-1	Sequence 1, Appli
42	c	39.8	3.3	319608	4	US-09-679-409-1	Sequence 1, Appli
43	c	39.6	3.2	355	4	US-09-621-976-15488	Sequence 15488, A
44		39.6	3.2	6583	4	US-10-204-708-25	Sequence 25, Appl
45		39.6	3.2	11131	4	US-10-204-708-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.8%; Score 58.4; DB 4; Length 1141;
Best Local Similarity 11.9%; Pred. No. 0.00022;
Matches 83; Conservative 261; Mismatches 352; Indels 4; Gaps 1;

```
Qy      518 AAAATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAG 577
            :::| :  :::: :|       | : : ::::::|: :::    |  :::: ::      |
Db      129 RWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTKYHMWNNNG 188

Qy      578 GAAGCACAATCAACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACA 637
              : :::::  : ::: ::  : :  :      :|:::::::::|:: :: :  |:  :
Db      189 CBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRM 248

Qy      638 CGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTT 697
                |   |: :     :|: : :  |       | |  ::|  : : : ::  ||||
Db      249 RNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNAKAMCKRAKYWGWNRABVNSTCTT 308

Qy      698 AGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAA 757
            ::::| : :|  :       |:| :::: :: :    :::        | ::| :    ::
Db      309 WKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNNWTYKKARHBARWDW 368

Qy      758 CTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTAT 817
            ::: |: :::   |:::   :|:::  | :    ||  :: : ::::::: : : |:
Db      369 VWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNN 428

Qy      818 TCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTT 877
                :|: |:           ::|   :::|::::        :: : |::  | ::  |
```

```
Db        429 NNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMT 488

Qy        878 CGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTAT 937

Db        489 NGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWNNYAAWYTKSS 548

Qy        938 TATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAG 997

Db        549 WNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTWBHBRAGAA 608

Qy        998 ATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCA 1057

Db        609 HYWMBMMYBAKCHCMKAWY----KAKKYAGAGGSNNNNNNNNNNNNNNNNNATCARDDYYA 664

Qy       1058 TGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCA 1117

Db        665 ASRWYAMANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNAGTWKNNN 724

Qy       1118 TATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACT 1177

Db        725 NNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAYTKYTTTN 784

Qy       1178 TTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAA 1217

Db        785 NNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNGWSDMWV 824
```

RESULT 2
US-09-806-708B-22/c
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match 4.5%; Score 54.4; DB 4; Length 1141;
Best Local Similarity 11.8%; Pred. No. 0.0018;
Matches 127; Conservative 374; Mismatches 566; Indels 11; Gaps 2;

```
Qy          16 TTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAATTGTTCTTGAAAATA 75

Db        1124 TYKKANNNNNNNNGMGKDWNRMDATKWSATGTAWWTNHAKRGATMCWYWYWTGTNRRWCM 1065

Qy          76 AACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTGAATATGACTGCACTG 135

Db        1064 RTYAMRTWYTRSNANWSCATKBMWWTMKWYATKYRTAWYAMWCAWRNNNMWCATNGYAKS 1005

Qy         136 AATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCAATTTGTTTATTGATG 195

Db        1004 CATNNAMWYATTRWAAYAAAKWARWAGNNMRMYGAAAGNKWGCMAAMATMGBWWADTAGK 945

Qy         196 AGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGTAGGCTTTGAAAATTT 255

Db         944 MCNNNNNNWTTDVRRMAMKAKNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRT 885

Qy         256 TAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAATCTTCAAAGAGATTCT 315
```

```
Db        884 RHHTRTCRRTKYNNNNNNARTVYWYHHAARRWMNAWWTRTNNNNNNNNNNACRNTRTWWA 825

Qy        316 CAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGCCTATCGAAATTGGTT 375

Db        824 BWKHSWCNNNNNNNNNNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMHAAVTT 765

Qy        376 TATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTCAAAAGTGGTATTTCT 435

Db        764 THTDWCYKTWMNTWYWDMMTTMBTTTTRNMTTSTNMTNNNNNNMWACTNNNNNNNMWKAYY 705

Qy        436 CTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTGGTCACTCTATGAACC 495

Db        704 AHATNNWGCWWNNTDARRTNNTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNN 645

Qy        496 ACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTT 555

Db        644 NNNNNNSCCTCTRMMTMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCR 585

Qy        556 TCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGTTCACAAACAACATTC 615

Db        584 DVTYTRNNTYCKSYAHSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAG 525

Qy        616 TAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAA 675

Db        524 TMWWRHWNNNNTDTRYYWWWKRWARBT-----TTVYDSMCNAKSMWRGNNWRAMKMWWAA 470

Qy        676 AGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCT 735

Db        469 NNDAGAMDHWTYWMGNNTMWMRRAWKMMNMAWCRRAYCCNNNNNRACVWHKHKMWRWTWK 410

Qy        736 TGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTG 795

Db        409 YMWKAACNNNNBKAMYMRVAWMMYSRDTTNTDWMMWTSDWBWHWYTVDYTMMRAWNNNNN 350

Qy        796 CTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAA 855

Db        349 NNWRBCKTTSWMWWMDHMNTHCTYGNNTWGSAYB------MAAMSMWAAGASNBVTYNWC 296

Qy        856 GAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCC 915

Db        295 WRMTYMGKTMTNNNNNNKAWYYRTKTVAWCNNRYYYDTAVWTBKRNYKYCYAYBWYYBMY 236

Qy        916 ACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTT 975

Db        235 MGKHHWBWWRRABHRSWNMWWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHH 176

Qy        976 TCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAA 1035

Db        175 WCATNNNMMWWWYAYMHHMHKKGKAAWTNNKTABRDDHBAHVKTYWYWRYDYWCAMCWMN 116

Qy       1036 AGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAA 1093

Db        115 AKAKVRTAMKHMWYYTDRYVSANNTGVRWMMMRWCMWWYSMNNRWYYRMGRKYTWAWM 58
```

RESULT 3
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

```
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             4.3%;  Score 52;  DB 4;  Length 640681;
  Best Local Similarity   54.8%;  Pred. No. 0.04;
  Matches  103;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy       1031 ATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTG 1090
Db      95585 AGTAAAAAAATATTAATATACTATCAATAATTGAAATATTTTTTAAAAATCTTTTTCTAT 95644

Qy       1091 AAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAA 1150
Db      95645 TTAAAAAACATATTTTTTATATATCTGACTAGTATTTTTTAAAAAAATCATTTTAATATT 95704

Qy       1151 TGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAA 1210
Db      95705 TTTTTGATATTAAAATGATTTTTTATTCTTAATATATATTATTCAGTTTATATAAAATAA 95764

Qy       1211 AAAAAAAA 1218
Db      95765 AAAATATA 95772

RESULT 4
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT:  Jobert, S.
;  APPLICANT:  Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE:  2000-07-21
;  NUMBER OF SEQ ID NOS:  19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

  Query Match             3.9%;  Score 47.8;  DB 4;  Length 832;
  Best Local Similarity   17.4%;  Pred. No. 0.054;
  Matches   78;  Conservative  169;  Mismatches  201;  Indels    1;  Gaps    1;

Qy        692 CATCTTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATT 751
Db          1 YRWYWWKYTTWYAKCWTKWKWSWSYWMYWKWYYMKTYWRWRRKKKKAWWKYWKTWTWWYW 60

Qy        752 ATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGC 811
Db         61 RYAMWGTYKKKAMCRTKTKKKKKKGYMWMWYWGWRRSYMAMWTRTWTGYAYYRSMMYWWR 120

Qy        812 TTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACC 871
Db        121 YRCWKKKAYYRKTTCYSSKGWTWWKRWKKAWTTWWWKKTYYWAATRYWWMMCWTKRWRAS 180

Qy        872 ATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCT 931
Db        181 WWYCWWWGKARKWSTWRKSRSYASARSAKRCCYSCSWGAMSWKYMWRMWRWRGWATGAGM 240

Qy        932 ACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTT 991
Db        241 KAWRASCMMRRKYAGKSKTSYKSMWMCWTRSWKYCYTKARWTGYYCYRKGGMWGKRGRWY 300

Qy        992 CTACAGATTACCCAAAGCTCGTGTCATAATTG-CTGAAGAATTAAAGAAGTGGAACCTTA 1050
Db        301 ASKKYMWKRWWWCWARMYRYSTGTRASMWWRRWYYTMMMKWWKYAWARAAWRWWAMWWAW 360

Qy       1051 AGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTT 1110
Db        361 RRACAAAATATAATTATTTATGGTACAATTCTTGTACTTTAGCAAATCTGGAGTTAGTTC 420

Qy       1111 AAATGCATATTTTATTAGTAATACTAACA 1139
Db        421 ATAGTCAAAGTCAGTTAATATTTCTTAGA 449

RESULT 5
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match             3.7%;  Score 45.6;  DB 4;  Length 640681;
  Best Local Similarity   44.6%;  Pred. No. 1.2;
  Matches  267;  Conservative    0;  Mismatches  324;  Indels    7;  Gaps    2;

Qy        622 ACGAAGATATCTAACACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTAAAGTGGA 681
Db     455530 AGGGATAAATATAATGCAATTAAAAATTTTTTTATTTAAATTATTATTTTAATATTTTT 455471

Qy        682 AATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACT 741
Db     455470 TTCGTTTTCTTTTAATAATTTTGCATTCGGATTTTTTTTAATAATCGATATGATTCTAA 455411

Qy        742 GGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTT 801
Db     455410 TTTTTTTATTGATAAAAAATATACAAAAGATTTATTTTTTTAAAAAAAATGAATGTTT 455351

Qy        802 TCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAA 861
Db     455350 TTCTTTTTTAAAATATAAAAAAAATTTTTTTTAAATAAAAAAATTATTATTGGCAATAA 455291

Qy        862 TCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACA 921
Db     455290 TCATTTTATTGGTTTTTTTCAAAAAAATAGATTTAAG-ATATTTTATATTGTGAAATCAA 455232

Qy        922 TATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCAT 981
Db     455231 AAGACACGATGTATTCTATTGCTAAAAATTCTGGTTATAATTATCATGAATTATCTAAAT 455172

Qy        982 AAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGT 1041
Db     455171 TTAACTCTATCAAAAAACCTTATAAA------ATAATTATAGGTCAAAAAATATGGATGG 455118

Qy       1042 GGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAA 1101
Db     455117 GAGATTTTTTAATTGATAAAAATAATAATGATTGTTCTATTCTAAATTTAGAAAAAAACA 455058

Qy       1102 ATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGT 1161
```

```
                | ||    ||    ||  ||    |||  |  | | || | ||  |||| |   |   |
Db     455057 GTATAAAACAACATAATTCTTGTGAAGTTGTTTTTAAAAATTTATTAAATATTGAAAAAT 454998

Qy       1162 TTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAAAA 1219
              || |  |       |    | |   | || |   | ||      | ||||||||| |||
Db     454997 TTTTGAAAGATAATATAAAAACTAAAAAAATATGTTTTTTTTTGCATAAAAAAAAATAAA 454940


RESULT 6
US-09-614-912-35
; Sequence 35, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
;  APPLICANT:  Allen, Steve
;  APPLICANT:  Rafalski, Antoni
;  APPLICANT:  Orozco, Buddy
;  APPLICANT:  Miao, Gou-Hau
;  APPLICANT:  Famodu, Omolayo O.
;  APPLICANT:  Lee, Jian Ming
;  APPLICANT:  Sakai, Hajime
;  APPLICANT:  Weng, Zude
;  APPLICANT:  Caimi, Perry G
;  APPLICANT:  Anderson, Shawn
;  TITLE OF INVENTION:  Plant Metabolism Genes
;  FILE REFERENCE:  BB1378 US NA
;  CURRENT APPLICATION NUMBER: US/09/614,912
;  CURRENT FILING DATE:  2000-07-12
;  PRIOR APPLICATION NUMBER:  60/143,401
;  PRIOR FILING DATE:  1999-07-12
;  PRIOR APPLICATION NUMBER:  60/143,412
;  PRIOR FILING DATE:  1999-07-12
;  PRIOR APPLICATION NUMBER:  60/146,650
;  PRIOR FILING DATE:  1999-07-30
;  PRIOR APPLICATION NUMBER:  60/170,906
;  PRIOR FILING DATE:  1999-12-15
;  PRIOR APPLICATION NUMBER:  60/172,959
;  PRIOR FILING DATE:  1999-12-21
;  PRIOR APPLICATION NUMBER:  60/172,946
;  PRIOR FILING DATE:  1999-12-21
;  NUMBER OF SEQ ID NOS:  204
;  SOFTWARE:  Microsoft Office 97
; SEQ ID NO 35
;   LENGTH: 1436
;   TYPE: DNA
;   ORGANISM: Zea mays
US-09-614-912-35

  Query Match             3.7%;  Score 45.4;  DB 4;  Length 1436;
  Best Local Similarity   65.0%;  Pred. No. 0.22;
  Matches   67;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy       1117 ATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTAC 1176
              ||| | || || ||||| ||| |     || ||  ||  ||| || |  ||| | ||||
Db       1330 ATAGTATAATAATAATAATAAAATAGCAAGTAATAGTTGTATCGTATTATTATTAATTAA 1389

Qy       1177 TTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219
              ||| | | |||  |    |||  ||  ||||||||||||||||
Db       1390 TTTATTAGCTGGTAGGCAAGTAGTATTAAAAAAAAAAAAAAAA 1432


RESULT 7
US-10-204-708-22
; Sequence 22, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT:  PIEPENBROCK, Christian
;  APPLICANT:  BERLIN, Kurt
;  TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
;  TITLE OF INVENTION:  by Assessing DNA Methylation
;  FILE REFERENCE: 5013.1012
;  CURRENT APPLICATION NUMBER: US/10/204,708
;  CURRENT FILING DATE:  2003-05-06
;  PRIOR APPLICATION NUMBER: PCT/EP01/03971
;  PRIOR FILING DATE: 2001-04-06
;  PRIOR APPLICATION NUMBER: DE 10019058.8
;  PRIOR FILING DATE: 2000-04-06
;  PRIOR APPLICATION NUMBER: DE 10019173.8
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: DE 10032529.7
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: DE 10043826.1
;  PRIOR FILING DATE: 2000-09-01
;  NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 22
;   LENGTH: 11049
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-22

  Query Match             3.6%;  Score 44;  DB 4;  Length 11049;
  Best Local Similarity   44.7%;  Pred. No. 0.85;
  Matches  170;  Conservative    0;  Mismatches  210;  Indels    0;  Gaps    0;

Qy         12 TAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAATTGTTCTTGAA 71
              ||| || |||| |   | || ||  || ||| | |   |||  |||    | | | | ||
Db       3371 TAATTTGAAAATATTTGTATAATTTGATTTGTTTAATTTTAGTAAGTTTATATACGTAAA 3430

Qy         72 AATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTGAATATGACTGC 131
              | ||||     |   |  | | | |||||| |    |||  |       |  | |
Db       3431 ATTAAATTTTTGCGGATAGAAAAGTTAAAATTTGTAGTAAAGTTAAAAATTATTTTCGGT 3490

Qy        132 ACTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCAATTTGTTTATT 191
              |    || |    | | | |   |   |          ||| ||  |  | | || |
Db       3491 ATAAGATTTAAAGATTAATAATATTTTAATATAAAAAATAAGTTTTTTTAGTTTTTTAGT 3550

Qy        192 GATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGTAGGCTTTGAAA 251
              | | ||  ||| | ||      |   |  | |          |    |   ||||  |
Db       3551 TAGGTGATTTAAAAATTTTTATTAGTATATATAGTTTAAAGAGATGGTAATATTTTGTTA 3610

Qy        252 ATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAATCTTCAAAGAGA 311
              || |    |  | |   | | |        ||| | | | ||||   ||  || |   |
Db       3611 ATATATTTAAACGTATATTATAAATATTAAAATTTATAAGAGATAAAAGTTTAATTTATT 3670

Qy        312 TTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGCCTATCGAAATT 371
              ||   | ||||  | | | | | || ||  | |||    |  | | |    |||  | ||
Db       3671 TTTGTATTATTTGATATAGATATTTAGATTGTATAAATAAGATATTAAGTTATATAGATG 3730

Qy        372 GGTTTATTCTTAACTACCTT 391
              | || |   |||| ||  ||
Db       3731 GTTTAAAGATTAAATAATTT 3750


RESULT 8
US-09-412-554A-3
; Sequence 3, Application US/09412554A
; Patent No. 6355788
; GENERAL INFORMATION:
;  APPLICANT:  Conklin, Darrell
;  APPLICANT:  Ellsworth, Jeff L.
;  TITLE OF INVENTION:  FOLLISTATIN RELATED PROTEIN ZFSTA2
;  FILE REFERENCE:  98-50
;  CURRENT APPLICATION NUMBER: US/09/412,554A
;  CURRENT FILING DATE:  1999-10-05
;  NUMBER OF SEQ ID NOS:  15
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 2949
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Degenerate oligonucleotide sequence encoding the
;   OTHER INFORMATION: zfsta2 polypeptide of SEQ ID NO:2.
;   NAME/KEY: variation
```

```
;   LOCATION: (1)...(2949)
;   OTHER INFORMATION: Each N is independently any nucleotide.
US-09-412-554A-3

  Query Match             3.6%;  Score 43.8;  DB 3;  Length 2949;
  Best Local Similarity   30.3%;  Pred. No. 0.64;
  Matches   82;  Conservative   51;  Mismatches  138;  Indels    0;  Gaps    0;

Qy         73 ATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTGAATATGACTGCA 132
              |:||:|| |    :   :  ||      |:||:  : | ||  |:  :||:     |: |
Db        290 AYAARCCNGTNTGYGGNWSNGAYGGNGARTTYTAYGARAAYCAYTGYGARGTNCAYMGNG 349

Qy        133 CTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCAATTTGTTTATTG 192
              | |  | :   ||:  : | ||:  : |   :|  | :||:     :  :|| ||:|  |
Db        350 CNGCNTGYYTNAARAARCARAARATHACNATHGTNCAYAAYGARGAYTGYTTYTTYAARG 409

Qy        193 ATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGTAGGCTTTGAAAA 252
              || ||:|   |    || ||:||::  | :|       |:      :  | ::|  |:|
Db        410 GNGAYAARTGYAARACNACNGARTAYWSNAARATGAARAAYATGYTNYTNGAYYTNCARA 469

Qy        253 TTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAATCTTCAAAGAGAT 312
               :  :||:| : |:    |:     :| :||     | |    :||   ::: :  | :
Db        470 AYCARAARTAYATHATGCARGARAAYGARAAYCCNAAYGGNGAYGAYATHWSNMGNAARA 529

Qy        313 TCTCAATATTAAAGAAAAAGCTTAAGCATTT 343
              :   |  |  |  |:| | |:||: |:||
Db        530 ARYTNYTNGTNGAYCARATGTTYAARTAYTT 560


RESULT 9
US-08-947-823-1/c
; Sequence 1, Application US/08947823
; Patent No. 6114605
;  GENERAL INFORMATION:
;    APPLICANT:  Williamson, Valerie M.
;    APPLICANT:  Kaloshian, Isgouhi
;    APPLICANT:  Yaghoobi, Jafar
;    APPLICANT:  Bodeau, John
;    APPLICANT:  Milligan, Stephen
;    TITLE OF INVENTION:  Procedures and Materials for Conferring
;    TITLE OF INVENTION:  Pest Resistance in Plants
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/947,823
;      FILING DATE:  09-OCT-1997
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/18802
;      FILING DATE:  09-OCT-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/028,191
;      FILING DATE:  10-OCT-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  023070-070210US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  51952 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-947-823-1

  Query Match             3.6%;  Score 43.8;  DB 3;  Length 51952;
  Best Local Similarity   50.2%;  Pred. No. 1.5;
  Matches  108;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy        746 ATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTT 805
              ||||||||  |  || || || || ||  ||| | ||  |  |||  |  |   ||
Db      12778 ATTATTATATACATATAAATTATAGATGATAAATATTAATAGATTAGATATTAAATTAGG 12719

Qy        806 CAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCAC 865
              | |  || ||        |  || |||| || ||  ||| |||||  |  |  ||  |
Db      12718 AATTAATTTTAGGAAGTCTAAAAGTCATTAACATTTAATTAAATTTATATAATCAAATAT 12659

Qy        866 TTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATT 925
              || |  |||||   |  ||  |||   |   |  || ||  |||   | ||       |
Db      12658 TTAAATATTTTATAACAATTTAATTTATAGTAAAAGTAAAAACGTAATTCAACTTTCAAC 12599

Qy        926 CTCTCTACTTATTATGGGTGGTATGTAATATTAGA 960
              | | || | ||| |    | ||| ||    || |
Db      12598 ATTTTTATTCATTCTTTTAGTTATATATATATATA 12564


RESULT 10
US-09-270-767-29614/c
; Sequence 29614, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29614
;   LENGTH: 826
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-29614

  Query Match             3.6%;  Score 43.4;  DB 4;  Length 826;
  Best Local Similarity   48.6%;  Pred. No. 0.55;
  Matches  119;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy        948 ATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAA 1007
              ||  |||| |  | | ||  |  |   |||  |  | |   ||| || | || || ||
Db        482 ATTAAATAGTTTATTCAATTTGTCAAGTTCTTAAGATTGTGTTTTTAGAAATGACACATC 423

Qy       1008 GCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCT 1067
              |  | | |   |      |   |||  |||| || |   |  | |||||  |  |  ||
Db        422 TATTTTATAAACCTCTGCAAGTTATTCGAGAAATGCATTTTAGAAATTCAAAATCGGTCT 363

Qy       1068 ATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTA 1127
              |||||  |  ||    ||  | | ||| || |  | ||| ||   ||  |  | | |||  |
Db        362 TTTTTTTAGCAACAGCAATTCTTAAATTATACTACTTTTTTCCCAAATCGAAATATATAA 303

Qy       1128 GTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTG 1187
              |  |    | | |||||      || |  |  |  ||     ||||  |||  | ||
Db        302 CTTTTGTCTATATTTGTATATTGAGTATATTTCCTCTTTGTAAGATTAACATTTTGTTTG 243

Qy       1188 AGAAA 1192
              | |
Db        242 TTACA 238


RESULT 11
US-09-270-767-13608/c
```

```
; Sequence 13608, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13608
;   LENGTH: 1348
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-13608

  Query Match             3.6%;  Score 43.4;  DB 4;  Length 1348;
  Best Local Similarity   48.6%;  Pred. No. 0.63;
  Matches  119;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy        948 ATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAA 1007
              ||  |||| |  | | ||  |  |   |||  |  | |   ||| || | || || ||
Db        482 ATTAAATAGTTTATTCAATTTGTCAAGTTCTTAAGATTGTGTTTTTAGAAATGACACATC 423

Qy       1008 GCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCT 1067
                | | | |    |      |   |||  |||| || |   |  | |||||  |  |  ||
Db        422 TATTTTATAAACCTCTGCAAGTTATTCGAGAAATGCATTTTAGAAATTCAAAATCGGTCT 363

Qy       1068 ATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTA 1127
              |||||  |  ||   ||  | | ||| || |  | ||| ||   ||  |  | | ||| |
Db        362 TTTTTTTAGCAACAGCAATTCTTAAATTATACTACTTTTTTCCCAAATCGAAATATATAA 303

Qy       1128 GTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTG 1187
                |  |    | | |||||      || |  |   ||     ||||   |||  | ||
Db        302 CTTTTGTCTATATTTGTATATTGAGTATATTTCCTCTTTGTAAGATTAACATTTTGTTTG 243

Qy       1188 AGAAA 1192
                 | |
Db        242 TTACA 238


RESULT 12
US-09-710-279-3580/c
; Sequence 3580, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3580
;   LENGTH: 2999
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3580

  Query Match             3.6%;  Score 43.4;  DB 4;  Length 2999;
  Best Local Similarity   46.3%;  Pred. No. 0.8;
  Matches  143;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy        893 TGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTA 952
              | ||| | ||    |   |||| ||      ||| | |||| | || |        |  ||
Db        458 TACAAACGGATGGTTTACTTTGCATCTATAATTTTTTCTATTGATCAACCACTATTCGTG 399

Qy        953 ATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCG 1012
                | | | | | |         ||  | |   |   ||||     | ||   | | |
Db        398 CGCTGATAGTAAGCTAACTACATTAACCTTTTTATTTTTCCTGCAAATAATTATAAATAC 339

Qy       1013 TGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTT 1072
              | |  ||          |  || |  | |     || |  ||| || | | |  || | |
Db        338 TCTTCTATAAAAATTTAATCTACATCAATATTTTCTAATAATTAATTACGCATATAATGT 279

Qy       1073 TGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAAT 1132
              |||||  |  | |  |||  | ||  |||  || | || | ||     | || ||||   ||
Db        278 TGAAACACTTTATATTTCCATATCCATATTTTAAATTCATATTATAAATTAATTAAAGAT 219

Qy       1133 ACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAA 1192
                | | |||| |||  |||   |       | ||||  | ||   ||  | ||    |||
Db        218 CGTGATAATTTTAGTTAATACATATACTGCTTTTTATATTTTTTATTCAATTCATGAAAA 159

Qy       1193 ACAGTCTTA 1201
                |   | |||
Db        158 AATTTTTTA 150


RESULT 13
US-09-710-279-4201/c
; Sequence 4201, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4201
;   LENGTH: 3099
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4201

  Query Match             3.6%;  Score 43.4;  DB 4;  Length 3099;
  Best Local Similarity   46.3%;  Pred. No. 0.8;
  Matches  143;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy        893 TGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTA 952
              | ||| | ||    |   |||| ||      ||| | |||| | || |        |  ||
Db       3051 TACAAACGGATGGTTTACTTTGCATCTATAATTTTTTCTATTGATCAACCACTATTCGTG 2992

Qy        953 ATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCG 1012
                | | | | | |         ||  | |   |   ||||     | ||   | | |
Db       2991 CGCTGATAGTAAGCTAACTACATTAACCTTTTTATTTTTCCTGCAAATAATTATAAATAC 2932

Qy       1013 TGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTT 1072
              | |  ||          |  || |  | |     || |  ||| || | | |  || | |
Db       2931 TCTTCTATAAAAATTTAATCTACATCAATATTTTCTAATAATTAATTACGCATATAATGT 2872

Qy       1073 TGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAAT 1132
              |||||  |  | |  |||  | ||  |||  || | || | ||     | || ||||   ||
Db       2871 TGAAACACTTTATATTTCCATATCCATATTTTAAATTCATATTATAAATTAATTAAAGAT 2812

Qy       1133 ACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAA 1192
                | | |||| |||  |||   |       | ||||  | ||   ||  | ||    |||
Db       2811 CGTGATAATTTTAGTTAATACATATACTGCTTTTTATATTTTTTATTCAATTCATGAAAA 2752

Qy       1193 ACAGTCTTA 1201
                |   | |||
Db       2751 AATTTTTTA 2743


RESULT 14
```

```
US-09-710-279-3614
; Sequence 3614, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3614
;   LENGTH: 3501
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3614

  Query Match             3.6%;  Score 43.4;  DB 4;  Length 3501;
  Best Local Similarity   46.3%;  Pred. No. 0.83;
  Matches  143;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy        893 TGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTA 952
              | ||| | ||    |   |||| ||      ||| | |||| | || |       |  ||
Db       2611 TACAAACGGATGGTTTACTTTGCATCTATAATTTTTTCTATTGATCAACCACTATTCGTG 2670

Qy        953 ATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCG 1012
              | | | | | |         ||  | |   |   ||||     | ||  | |  |
Db       2671 CGCTGATAGTAAGCTAACTACATTAACCTTTTTATTTTTCCTGCAAATAATTATAAATAC 2730

Qy       1013 TGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTT 1072
              | |  ||          |  || |  | |     || |  ||| || | | |  || | |
Db       2731 TCTTCTATAAAAATTTAATCTACATCAATATTTTCTAATAATTAATTACGCATATAATGT 2790

Qy       1073 TGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAAT 1132
              |||||  |  | |  |||  | ||  |||  || | || | ||    | || ||||  ||
Db       2791 TGAAACACTTTATATTTCCATATCCATATTTTAAATTCATATTATAAATTAATTAAAGAT 2850

Qy       1133 ACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAA 1192
              | | |||| |||  |||   |       | ||||  | ||   ||  | ||    |||
Db       2851 CGTGATAATTTTAGTTAATACATATACTGCTTTTTATATTTTTTATTCAATTCATGAAAA 2910

Qy       1193 ACAGTCTTA 1201
              |   | |||
Db       2911 AATTTTTTA 2919


RESULT 15
US-09-710-279-3619/c
; Sequence 3619, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3619
;   LENGTH: 4015
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-3619

  Query Match             3.6%;  Score 43.4;  DB 4;  Length 4015;
  Best Local Similarity   46.3%;  Pred. No. 0.87;
  Matches  143;  Conservative    0;  Mismatches  166;  Indels    0;  Gaps    0;

Qy        893 TGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTA 952
              | ||| | ||    |   |||| ||      ||| | |||| | || |       |  ||
Db       1134 TACAAACGGATGGTTTACTTTGCATCTATAATTTTTTCTATTGATCAACCACTATTCGTG 1075

Qy        953 ATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCG 1012
              | | | | | |         ||  | |   |   ||||     | ||  | |  |
Db       1074 CGCTGATAGTAAGCTAACTACATTAACCTTTTTATTTTTCCTGCAAATAATTATAAATAC 1015

Qy       1013 TGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTT 1072
              | |  ||          |  || |  | |     || |  ||| || | | |  || | |
Db       1014 TCTTCTATAAAAATTTAATCTACATCAATATTTTCTAATAATTAATTACGCATATAATGT 955

Qy       1073 TGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAAT 1132
              |||||  |  | |  |||  | ||  |||  || | || | ||    | || ||||  ||
Db        954 TGAAACACTTTATATTTCCATATCCATATTTTAAATTCATATTATAAATTAATTAAAGAT 895

Qy       1133 ACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAA 1192
              | | |||| |||  |||   |       | ||||  | ||   ||  | ||    |||
Db        894 CGTGATAATTTTAGTTAATACATATACTGCTTTTTATATTTTTTATTCAATTCATGAAAA 835

Qy       1193 ACAGTCTTA 1201
              |   | |||
Db        834 AATTTTTTA 826


Search completed: January 19, 2005, 17:04:40
Job time : 378 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 18, 2005, 20:57:16 ; Search time 678 Seconds
(without alignments)
9438.118 Million cell updates/sec

Title: US-10-070-666A-1
Perfect score: 1219
Sequence: 1 agtaagcaaactaaatttaa.....taacaaaaaaaaaaaaaaaa 1219

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1219	100.0	1219	4	AAF61947	Aaf61947 T. thermo
	2	614.4	50.4	2492	4	AAF61948	Aaf61948 T. thermo
	3	64	5.3	2000	8	ADA71938	Ada71938 Rice gene
	4	59	4.8	8056	8	ABZ10246	Abz10246 Haematopo
c	5	58	4.8	6103	6	ABL33690	Abl33690 Human imm
c	6	57.4	4.7	8056	8	ABZ10100	Abz10100 Haematopo
c	7	54.6	4.5	8056	8	ABZ10246	Abz10246 Haematopo
	8	53.2	4.4	1501	8	ABZ10188	Abz10188 Haematopo
	9	53.2	4.4	1501	10	ADE84162	Ade84162 Human lym
c	10	52.6	4.3	1199	3	AAA02542	Aaa02542 Human col
	11	52	4.3	110000	6	ABA92787_0	Aba92787 Buchnera
c	12	51.8	4.2	424	8	ABX46053	Abx46053 Bovine ES
c	13	51.8	4.2	7319	6	ABL34044	Abl34044 Human imm
c	14	51.4	4.2	5728	6	ABL32101	Abl32101 Human imm
c	15	51.2	4.2	5750	4	AAS46709	Aas46709 Tumour su
c	16	51.2	4.2	5750	6	ABL34009	Abl34009 Human imm
	17	50.8	4.2	3549	3	AAA70223	Aaa70223 Plasmodiu
	18	50.4	4.1	13427	6	ABL33927	Abl33927 Human imm
c	19	50	4.1	8951	6	ABL32795	Abl32795 Human imm
	20	50	4.1	16439	6	ABL32887	Abl32887 Human imm
c	21	49.8	4.1	5442	6	ABL33969	Abl33969 Human imm
c	22	49.8	4.1	10328	6	ABL33544	Abl33544 Human imm
c	23	49.4	4.1	2000	6	ABZ15104	Abz15104 Arabidops
	24	49.4	4.1	2000	6	ABZ14950	Abz14950 Arabidops
c	25	49.4	4.1	7676	6	ABL34598	Abl34598 Human met
c	26	49.4	4.1	7676	6	ABL70409	Abl70409 Chemicall
	27	49	4.0	8056	8	ABZ10100	Abz10100 Haematopo
c	28	48.8	4.0	1830	6	ABL56243	Abl56243 AmEPV met
	29	48.8	4.0	32392	6	ABL56203	Abl56203 AmEPV gen
c	30	48.6	4.0	6129	6	ABL34582	Abl34582 Human met
	31	48	3.9	1501	8	ABZ10042	Abz10042 Haematopo
	32	48	3.9	1501	10	ADE84086	Ade84086 Human lym
	33	47.8	3.9	1348	4	AAH74716	Aah74716 Nucleotid
c	34	47.8	3.9	16724	6	ABL33091	Abl33091 Human imm
c	35	47.8	3.9	16724	6	ABL34537	Abl34537 Human met
c	36	47.8	3.9	16724	6	ABL70260	Abl70260 Chemicall
c	37	47.6	3.9	16724	6	ABL33090	Abl33090 Human imm
c	38	47.6	3.9	16724	6	ABL34536	Abl34536 Human met
c	39	47.6	3.9	16724	6	ABL70259	Abl70259 Chemicall
c	40	47.4	3.9	17421	4	AAS45349	Aas45349 Chemicall
c	41	47.4	3.9	17421	6	ABK28182	Abk28182 DNA trans
c	42	47.4	3.9	19634	8	ABZ10162	Abz10162 Haematopo
c	43	47.2	3.9	6749	4	AAS46525	Aas46525 Tumour su
c	44	47.2	3.9	17703	6	ABK39953	Abk39953 Human che
c	45	47	3.9	2000	8	ADA71938	Ada71938 Rice gene

ALIGNMENTS

RESULT 1
AAF61947
ID AAF61947 standard; cDNA; 1219 BP.
XX
AC AAF61947;
XX
DT 22-AUG-2001 (first entry).
XX
DE T. thermophila delta-6-desaturase cDNA.
XX
KW Delta-6-desaturase; antirheumatic; antiarthritic; antiarteriosclerotic;
KW antidiabetic; neuroleptic; dermatological; gene therapy; ciliate;
KW delta-6-unsaturated fatty acid production; gamma-linolenic acid; GLA;
KW eicosanoid biosynthesis; prostaglandin; atopic eczema; atherosclerosis;
KW rheumatoid arthritis; diabetic neuropathy; schizophrenia; cosmetic;
KW animal nutrition; human nutrition; stearidonic acid;
KW plant transformation; oilseed crop; ss.
XX
OS Tetrahymena thermophila.
XX
FH Key Location/Qualifiers
FT CDS 11. .1091
FT /*tag= a
FT /product= "delta-6-desaturase"
FT /note= "contains in-frame stop codons"
FT /transl_except= (pos:51. .53, aa:Gln)
FT /transl_except= (pos:201. .203, aa:Gln)
FT /transl_except= (pos:240. .242, aa:Gln)
FT /transl_except= (pos:285. .287, aa:Gln)
FT /transl_except= (pos:633. .635, aa:Gln)
FT /transl_except= (pos:654. .656, aa:Gln)
FT /transl_except= (pos:759. .761, aa:Gln)
FT /transl_except= (pos:951. .953, aa:Gln)
FT /transl_except= (pos:957. .959, aa:Gln)
FT /transl_except= (pos:981. .983, aa:Gln)
XX
PN DE10044468-A1.
XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000DE-01044468.
XX
PR 10-SEP-1999; 99DE-01043270.
XX

PA (AXIV-) AXIVA GMBH.
XX
DR WPI; 2001-245832/26.
DR P-PSDB; AAB70934.
XX
PT New nucleic acid encoding delta-6-desaturase, useful for producing
PT ciliates and plants that overproduce unsaturated fatty acids, derived
PT from Tetrahymena.
XX
PS Claim 1; Page 13; 30pp; German.
XX
CC This invention describes a novel nucleic acid (I) encoding a delta-6-
CC desaturase from Tetrahymena thermophila. The products of the invention
CC have antirheumatic, antiarthritic, antiarteriosclerotic, antidiabetic,
CC neuroleptic and dermatological and can be used for gene therapy. (I) is
CC overexpressed in ciliates to increase production of delta-6-unsaturated
CC fatty acids, specifically gamma-linolenic acid (GLA) which is an
CC intermediate in biosynthesis of eicosanoids (and e.g. prostaglandins) and
CC useful for treating e.g. atopic eczema, rheumatoid arthritis,
CC atherosclerosis, diabetic neuropathy, schizophrenia and in cosmetics and
CC in animal or human nutrition. Also production of stearidonic acid is
CC increased. (I) may also be used to transform plants, particularly oilseed
CC crops, optionally in combination with other desaturases. Fragments of (I)
CC are useful for producing individual epitopes, and as probes for
CC identifying related sequences, as antisense sequences and as primers for
CC polymerase chain reaction. The desaturase, or its fragments, can be used
CC to raise antibodies, used e.g. to isolate homologous proteins.
CC Tetrahymena transformed with (I) are easy to grow on a large scale and at
CC high cell density, without being affected by weather and environmental
CC pollutants. They produce a pattern of fatty acids far simpler than that
CC produced by higher organisms. This cDNA sequence encodes the Tetrahymena
CC thermophila delta-6-desaturase protein described in the method of the
CC invention
XX
SQ Sequence 1219 BP; 420 A; 204 C; 182 G; 413 T; 0 U; 0 Other;

Query Match 100.0%; Score 1219; DB 4; Length 1219;
Best Local Similarity 100.0%; Pred. No. 1.8e-227;
Matches 1219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGTAAGCAAACTAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTAAGCAAACTAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAA 60

Qy         61 TTGTTCTTGAAAATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGTTCTTGAAAATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTG 120

Qy        121 AATATGACTGCACTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATATGACTGCACTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCA 180

Qy        181 ATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGT 240

Qy        241 AGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAAT 300

Qy        301 CTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGC 360

Qy        361 CTATCGAAATTGGTTTATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTATCGAAATTGGTTTATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTC 420

Qy        421 AAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTG 480

Qy        481 GTCACTCTATGAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTCACTCTATGAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTC 540

Qy        541 CTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGT 600

Qy        601 TCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCC 660

Qy        661 CCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAG 720

Qy        721 GAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTC 780

Qy        781 TTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATG 840

Qy        841 AAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCA 900

Qy        901 GAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGA 960

Qy        961 CTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAA 1020

Qy       1021 TTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAAT 1080

Qy       1081 CTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAA 1140

Qy       1141 TTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTT 1200

Qy       1201 AACAAAAAAAAAAAAAAAA 1219
              ||||||||||||||||||||
Db       1201 AACAAAAAAAAAAAAAAAA 1219
```

RESULT 2
AAF61948
ID AAF61948 standard; DNA; 2492 BP.
XX
AC AAF61948;
XX
DT 22-AUG-2001 (first entry)
XX
DE T. thermophila delta-6-desaturase genomic DNA.
XX
KW Delta-6-desaturase; antirheumatic; antiarthritic; antiarteriosclerotic;
KW antidiabetic; neuroleptic; dermatological; gene therapy; ciliate;
KW delta-6-unsaturated fatty acid production; gamma-linolenic acid; GLA;
KW eicosanoid biosynthesis; prostaglandin; atopic eczema; atherosclerosis;
KW rheumatoid arthritis; diabetic neuropathy; schizophrenia; cosmetic;
KW animal nutrition; human nutrition; stearidonic acid;
KW plant transformation; oilseed crop; ds.
XX
OS Tetrahymena thermophila.
XX
PN DE10044468-A1.

XX
PD 15-MAR-2001.
XX
PF 08-SEP-2000; 2000DE-01044468.
XX
PR 10-SEP-1999; 99DE-01043270.
XX
PA (AXIV-) AXIVA GMBH.
XX
DR WPI; 2001-245832/26.
XX
PT New nucleic acid encoding delta-6-desaturase, useful for producing
PT ciliates and plants that overproduce unsaturated fatty acids, derived
PT from Tetrahymena.
XX
PS Example 6; Page 15-16; 30pp; German.
XX
CC This invention describes a novel nucleic acid (I) encoding a delta-6-
CC desaturase from Tetrahymena thermophila. The products of the invention
CC have antirheumatic, antiarthritic, antiarteriosclerotic, antidiabetic,
CC neuroleptic and dermatological and can be used for gene therapy. (I) is
CC overexpressed in ciliates to increase production of delta-6-unsaturated
CC fatty acids, specifically gamma-linolenic acid (GLA) which is an
CC intermediate in biosynthesis of eicosanoids (and e.g. prostaglandins) and
CC useful for treating e.g. atopic eczema, rheumatoid arthritis,
CC atherosclerosis, diabetic neuropathy, schizophrenia and in cosmetics and
CC in animal or human nutrition. Also production of stearidonic acid is
CC increased. (I) may also be used to transform plants, particularly oilseed
CC crops, optionally in combination with other desaturases. Fragments of (I)
CC are useful for producing individual epitopes, and as probes for
CC identifying related sequences, as antisense sequences and as primers for
CC polymerase chain reaction. The desaturase, or its fragments, can be used
CC to raise antibodies, used e.g. to isolate homologous proteins.
CC Tetrahymena transformed with (I) are easy to grow on a large scale and at
CC high cell density, without being affected by weather and environmental
CC pollutants. They produce a pattern of fatty acids far simpler than that
CC produced by higher organisms. This DNA sequence encodes the Tetrahymena
CC thermophila delta-6-desaturase protein described in the method of the
CC invention
XX
SQ Sequence 2492 BP; 952 A; 331 C; 309 G; 900 T; 0 U; 0 Other;

Query Match 50.4%; Score 614.4; DB 4; Length 2492;
Best Local Similarity 99.8%; Pred. No. 5.3e-110;
Matches 615; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       589 AACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAAT 648
            | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1652 AGCATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAAT 1711

Qy       649 TGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATT 708
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1712 TGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATT 1771

Qy       709 ATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCT 768
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1772 ATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCT 1831

Qy       769 ATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAA 828
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1832 ATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAA 1891

Qy       829 ATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAA 888
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1892 ATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAA 1951

Qy       889 TAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTA 948
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1952 TAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTA 2011

Qy       949 TGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAG 1008
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2012 TGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAG 2071

Qy       1009 CTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTA 1068
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2072 CTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTA 2131

Qy       1069 TTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAG 1128
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2132 TTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAG 2191

Qy       1129 TAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGA 1188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2192 TAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGA 2251

Qy       1189 GAAAACAGTCTTAACA 1204
              ||||||||||||||||
Db       2252 GAAAACAGTCTTAACA 2267
```

RESULT 3
ADA71938
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 5.3%; Score 64; DB 8; Length 2000;
Best Local Similarity 10.9%; Pred. No. 0.0043;
Matches 89; Conservative 362; Mismatches 357; Indels 8; Gaps 2;

```
Qy       103 TTATTTACAAGGATACTGAATATGACTGCACTGAATATGCTAAATCAAATAAGCATCCTG 162
```

```
Db        220 KYGTYKKMTYYSASRCMRAYMTTSYSWACSSYTWCRSKRRSMMWKMMRKMRWSRSYGWYS 279

Qy        163 GCGGTCTTAATTTCCTCAATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCA 222

Db        280 WSYKMWMCTAYKKSYYSRWCYMYRGGGWRGATRYWGRGYMSRMAMMYKKMYWYRGYKGMK 339

Qy        223 GAACACTCCATTCTAAGTAGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAA 282

Db        340 RGWWAGRMMMRSMCRWSKACYYMRWRMWRMTRRRRWAKKSSRTSRRKKRKWCMRKRKYKR 399

Qy        283 AA---TAAGAGGAGACTGAATCTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGC 339

Db        400 MRGYSRMRSCKRARWMKRCRSGRAWKMGCRGCMTCRMKSYGMMRWKSWKRMASKYKWMSR 459

Qy        340 ATTTATTCGAACCAAACTGGCCTATCGAAATTGGTTTATTCTTAACTACCTTTACTTTAT 399

Db        460 MYRWRKKKCSRTTMWGKTRGGMMGTMGRCRYKKRSGMKRKCRRRRWGRMYRMRWKRYYMS 519

Qy        400 TTGTCACTGGATGTTTGACTCAAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTAATGC 459

Db        520 ARYTMRYCARKKYSYSAARKARCWYRGKGYYWAGMWMKRYKRMYMYKMMWWYKRKYSKCS 579

Qy        460 AAATCATCAGTGGTTGGATTGGTCACTCTATGAACCACAATCGTAACCCTATATTAAGAA 519

Db        580 WYCKMSYYASCMKSARKAGAKMCKRSKMSAWSKSMRSSRKCRKCASKRSSAKRYAMMGGM 639

Qy        520 AATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGA 579

Db        640 TSGSRMSRWKSYTCYWRKWGSMKSTCTWMYYMSKYTYAKYGSYWRYRYRAWCMYMWRWYY 699

Qy        580 AGCACAATCAACATC-----ATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTA 634

Db        700 YRYRSYMTYMAWYTSSTRMAMTGMKYSGRYWTSWYKYCKCSWKYRSMWYYWSWWWAKTWK 759

Qy        635 ACACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCAT 694

Db        760 MWRRYATRMMWMWYRYSMKWYTWCTMWGYWWYWWRTYMKMRYMWYKCTKTYWYWSATYWT 819

Qy        695 CTTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATT 754

Db        820 GTWAAWWMAKTKMRMGMTGAKTRGRARKARYWWKWATWCATKRWMTKGKGAKWAWTWMAK 879

Qy        755 CAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTC 814

Db        880 AWRKYYWSWMRAWYYYYKTRRTRYKTCWWKARWGSWAYWRMWWKGSAKMWWMWKGGRWGW 939

Qy        815 TATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATT 874

Db        940 TKYWYWYCTTWKMACGRATKYMCCAGWWAMYSYSWTRTYWMRTWRWMWASSRTAKRMARM 999

Qy        875 TTTCGAACATCAAATAGCTGCAAGCAGAAACTACGC 910

Db       1000 MWKTRAWSKSYARAYWKMAGCACCTACACACTACAC 1035
```

RESULT 4
ABZ10246
ID ABZ10246 standard; DNA; 8056 BP.
XX
AC ABZ10246;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #386.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 386; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

Query Match 4.8%; Score 59; DB 8; Length 8056;
Best Local Similarity 44.3%; Pred. No. 0.047;
Matches 483; Conservative 0; Mismatches 585; Indels 23; Gaps 5;

```
Qy        101 ATTTATTTACAAGGATACTGAATATGACTGCACTGAATATGCTAAATCAAATAAGCATCC 160

Db       2396 ATTTTTTGAAAATTAAAAAAAAAAATAATTTATTTATTTTTAAAAATTAATTTTTTTTTAT 2455

Qy        161 TGGCGGTCTTAATTTCCTCAATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTT 220

Db       2456 TTTTATTTATATATTTTTTATTATTTTGTTTTAAAATAATTAAAATATTAATGAAAAATA 2515

Qy        221 CAGA-----ACACTCCATTCTAAGTAGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACT 275

Db       2516 AAAAAAAATAAATTAAATTATAAAATAAATTAAAAAATTTAAAAATTTAAATAAAATTTT 2575

Qy        276 GGCGCAAAATAAGAGGAGACTGAATCTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTT 335

Db       2576 TAAAAAATAAAAAATAAATGAAAAAAAAAATTATAAATATAAATAATTTAAAAAAAAAAA 2635

Qy        336 AAGCATTTATTCGAACCAAACTGGCCTATCGAAATTGGTTTATTCTTAACTACCTTTACT 395

Db       2636 ATAAAAAAAAATGTTAAAATTTAAAATTTTAAAAATAATAAATTTTTAAATTTTTTTAAA 2695

Qy        396 TTATTTGTCACTGGATGTTTGACTCAAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTA 455
```

```
Db       2696 TTTATTATTTTTTTTTTTAATAAAAATAATTTGTTAAATAAAT-----------TATTA 2744

Qy        456 ATGCAAATCATCAGTGGTTGGATTGGTCACTCTATGAACCACAATCGTAACCCTATATTA 515

Db       2745 ATAAAAATAAAAAAAATATTTAATTGTTTATTTTTAATAAATTTAAATGAATTTTGTTTA 2804

Qy        516 AGAAAATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGT 575

Db       2805 TTAAAATTGTTATTAATAATAAATGATATTTTTTTATAAAATATTAAAAAATTTGAAATTA 2864

Qy        576 AGGAAGCACAATCAACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAA 635

Db       2865 TTTTATGAAAAAAATTTTGTTGAAAAAAATAAAAAATTTGAAAATGAAAAAAAAATGTTA 2924

Qy        636 CACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATC 695

Db       2925 AAATTTTATGAAATGTAAAATTTAAATGAATTATTATTGTTAAAAAAATAAAAAAAAATT 2984

Qy        696 TTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTT---GCACTGGGTATTATTA 752

Db       2985 ATTGAAAAATTGATATATTTTAAAATGATGATGTTTAAATTTTATAAATATAAATTAAAA 3044

Qy        753 TTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCT 812

Db       3045 ATAAAATAAATGTTATAATTGTGAATGAAATTTGAATT-ATTGTTATTTTTATGTGTAAT 3103

Qy        813 TCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCA 872

Db       3104 TTAATAAAAAATTAATTTTTTTAAAATGAATTTAATGATGATTGAAAAATTGTTTTATAT 3163

Qy        873 TTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTA 932

Db       3164 ATAATTTTAAATAATAATATTTTTAATAAAAATTTTGTTGT---TTGAATTGTAAATAAA 3220

Qy        933 CTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTC 992

Db       3221 ATTTTAATTTATTATATTTTTTTTTAATAATTTGATTATTATTTTTAATTAATTGATTGT 3280

Qy        993 TACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAG 1052

Db       3281 AAAAAAATTTTTGTTATTGTTGTTGTTATTTTTTTGAAAAAAAAAAAATATAATTGTAAT 3340

Qy       1053 ATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAA 1112

Db       3341 GTAAAAAAATGAATAAAATTTGAAATGATGATTAAAAAATAAAAATGAAATAAATAATAA 3400

Qy       1113 ATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTA 1172

Db       3401 TTTGAATTTTTTTTAAATATTTTTGTTAAATGTTGAAAATTATAATTTGTTTTATTATTA 3460

Qy       1173 TTACTTTTTAA 1183

Db       3461 TTATTAATAAA 3471
```

RESULT 5
ABL33690/c
ID   ABL33690 standard; DNA; 6103 BP.
XX
AC   ABL33690;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1663.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 1663; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 6103 BP; 1794 A; 39 C; 1134 G; 3136 T; 0 U; 0 Other;

Query Match 4.8%; Score 58; DB 6; Length 6103;
Best Local Similarity 46.6%; Pred. No. 0.071;
Matches 221; Conservative 0; Mismatches 250; Indels 3; Gaps 1;

```
Qy        748 TATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCA 807

Db       1844 TCTACTTAAAACAAAATCTAACTACCTTTAAAACATCTCTACAAAATCCTTTACTCCTAA 1785

Qy        808 GTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAAT---GAAATTCGAAAGAAGAATCA 864

Db       1784 AAACTACTAATTTTATCCAAAACAAAATTAACCAACTATAAAAATCAAAACCAAAAAAAA 1725

Qy        865 CTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATAT 924

Db       1724 ACAACTCCTTTTCAAATCACACAACAAATTAAAAACTAAACAAAAATTATAACTTAAATC 1665

Qy        925 TCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAA 984

Db       1664 TTTTACTACATCGTATAATAACTTATTACATATCATCTCCACCACACCCTCTTTAATAAA 1605

Qy        985 TTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGA 1044

Db       1604 TCCCATAAAACAAAATACTCCAATCTCCAACTCATTAAAATACATAAAAAAAAAACAATA 1545

Qy       1045 ACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATT 1104

Db       1544 AAAAAAATAAACACTTTTTAACAAATAATTATAAAAAAAAACCTTTCCATACAAACATAT 1485

Qy       1105 TATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTG 1164

Db       1484 AATAATAAATACATCTATTATAAAAAATCCTACATAAAAAAAAAAATTAAATATACTAAA 1425

Qy       1165 TTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAA 1218

Db       1424 TACGTTCATTTTTCTAATCAAAAATTAAAACATATTAAAAAATAAAAACAAAAA 1371
```

RESULT 6

```
ABZ10100/c
ID   ABZ10100 standard; DNA; 8056 BP.
XX
AC   ABZ10100;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #240.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T,  Pelet C;
PI   Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
PT   disorders, comprises contacting a target nucleic acid with a reagent that
PT   distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS   Claim 28; SEQ ID NO 240; 117pp; English.
XX
CC   The present invention describes a method for detecting and
CC   differentiating between haematopoietic cell proliferative disorders
CC   associated with at least 1 gene and/or their regulatory regions in a
CC   subject. The method comprises contacting a target nucleic acid in a
CC   biological sample obtained from the subject with at least 1 reagent,
CC   which distinguishes between methylated and non-methylated CpG
CC   dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC   represent specifically claimed nucleotide sequences from the present
CC   invention. Oligonucleotides from the present invention can be used: for
CC   differentiating between healthy haematopoietic cells and proliferative
CC   disorder haematopoietic cells; for differentiating between acute
CC   lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC   determining the cytosine methylation state and/or single nucleotide
CC   polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC   related sequences and their complements; and as primers for the
CC   amplification of haematopoietic cell proliferation disorder related DNA
CC   sequences. The nucleotide sequences from the present invention can also
CC   be used for detecting a predisposition to, differentiation between
CC   subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC   haematopoietic cell proliferative disorders. The present method enables a
CC   highly specific classification of haematopoietic cell proliferative
CC   disorders allowing for improved and informed treatment of patients
XX
SQ   Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;

  Query Match             4.7%;  Score 57.4;  DB 8;  Length 8056;
  Best Local Similarity   44.3%;  Pred. No. 0.096;
  Matches  283;  Conservative    0;  Mismatches  351;  Indels    5;  Gaps    1;

Qy        586 ATCAACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACA 645
              || || |   ||| |    ||  |   || | ||  |  || |    |||    |  | |
Db       2746 ATTAATAATTTATTTAACGAATTATTTTTATTAAAAAAAAAAAAATAATAAATTTAAAAAA 2687

Qy        646 AATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTT 705
              | ||       |     || | |  ||| || |   ||      |   | ||   || ||

Db       2686 ATTTAAAAATTTATTATTTTTAAAATTTTAAATTTTAACGTTTTTTTTTATTTTTTTTT 2627

Qy        706 ATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACT 765
              | |   ||||  |   ||  |  ||   ||     |   | |||||| || ||  |
Db       2626 TTAAATTATTTATATTTATAATTTTTTTTTTCGTTTATTTTTTATTTTTTAAAAATTTTA 2567

Qy        766 TCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTG 825
              | || ||  | |   ||| | | ||  ||  |  |||  |   |  || | |  || | |
Db       2566 TTTAAATTTTTAAATTTTTTAATTTATTTTATAATTTAATTTATTTTTTTTTATTTTTCG 2507

Qy        826 GAAA-----TCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGA 880
                ||     | || |   | ||| |||||   ||| || |   |  | |  ||      |
Db       2506 TTAATATTTTAATTATTTTAAAACGAAATAATAAAAAATATATAAATAAAAATAAAAAAA 2447

Qy        881 ACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTAT 940
              |  | | |       |   | ||| ||   |||      |  |        || | |   |
Db       2446 AATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAATTTTCGAAAAATAATAAAATT 2387

Qy        941 GGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATT 1000
                |   | |      | |   || || |  |||    || || |    | ||   |
Db       2386 TAATAAATTATTTATAAAAATAAAAAATTATATTTAAAAATTAAATAAAATTTATTTAAA 2327

Qy       1001 ACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGA 1060
              | |||  |  |    || ||    || || ||||  |    |    |   ||
Db       2326 TACGAAAATTAATAAATTATTTTAAAAATAAATAAATTAAAAAATTAAAATTTTTTTAAT 2267

Qy       1061 AGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATAT 1120
                 ||  | | |||||    |  |||   || |   ||  ||| ||||   |  || |
Db       2266 TTATTTTAAATATTAAAAAATAAATTTTTATTAAAATTAAATTTTTTTTTTTTTTTATTT 2207

Qy       1121 TTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTT 1180
              |||  ||      |  | | | || ||  |||  | | ||  |||  |||  |||||  || |
Db       2206 TTTTTTTCGTTTTTTTAAATTTTTAATAAAATTATAATTTTTTTTTTTTTTTATTTTTTAT 2147

Qy       1181 TAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219
              |  |   | || | |   |||  ||||  | |  ||| |
Db       2146 TTTTAATAAAATAAAAAATTATTAAAATTATATTAAATA 2108


RESULT 7
ABZ10246/c
ID   ABZ10246 standard; DNA; 8056 BP.
XX
AC   ABZ10246;
XX
DT   16-JAN-2003  (first entry)
XX
DE   Haematopoietic cell proliferation disorder related DNA sequence #386.
XX
KW   Human; haematopoietic cell proliferation disorder; cytostatic;
KW   gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW   cytosine methylation state; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200277272-A2.
XX
PD   03-OCT-2002.
XX
PF   26-MAR-2002; 2002WO-EP003401.
XX
PR   26-MAR-2001; 2001US-0278333P.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Berlin K,  Braun A,  Distler J,  Guetig D,  Howe A,  Mueller J;
PI   Olek A,  Piepenbrock C,  Adorjan P,  Grabs G,  Lesche R,  Leu E;
PI   Lewin A,  Lipscher E,  Maier S,  Model F,  Mueller V,  Otto T,  Pelet C;
PI   Schwope I,  Ziebarth H;
XX
DR   WPI; 2003-018942/01.
XX
PT   Detecting and differentiating between hematopoietic cell proliferative
```

PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 386; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;

Query Match 4.5%; Score 54.6; DB 8; Length 8056;
Best Local Similarity 45.3%; Pred. No. 0.33;
Matches 282; Conservative 0; Mismatches 334; Indels 7; Gaps 2;

```
Qy       596 TATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTA 655
             ||| |  | | |  | ||  || |  |    |  |||   | || |||||    ||
Db      6415 TATATATATATAATATATAATATATAATTTTTTTTAATAAATATAATAAAAATTTATTTA 6356

Qy       656 ATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATT 715
               | || ||    |  |  |  |||        |  |||  ||  |  | | || |
Db      6355 TAACACATTTTTTAAATATACATAATAATACATTATTAAATTATTTTAATTATTTTAAAT 6296

Qy       716 TGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGT 775
               |    | ||  |    ||  | |   | || |||||  |  | ||  || || ||  |
Db      6295 ATCATTTTTTTTTTTATATTTTTATTATTTTTTTTTTATTTATATTAATATTTTACATTTT 6236

Qy       776 AATTCTTTCTGAATT--GATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCAT 833
             ||        |  |||   |||  |  ||| || | |  || ||   || ||    |  |
Db      6235 AAACACACATTTATTAATATTTTTTTTTTTTTTATTTTTTTTAAAATTTTTAATTAAAAC 6176

Qy       834 GAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCT 893
             ||  |    |   | |   || || |     | ||   |  | |  |||| | ||| |
Db      6175 ACAATTTTTTTTTTAATTTTAACAAAACATTTTATAATTTAATACACACATAACATAACA 6116

Qy       894 GCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAA 953
             || || | | |||  |||| |    | |     |    | |||     |  | | ||
Db      6115 ATAAACACATAATACATTTTCAATATAAAAAATAATTATTTCATTTACAATAATTTATAT 6056

Qy       954 TATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGT 1013
             |  || | | ||  ||    |||    |  |||    ||  | | |       |||
Db      6055 TTATAAAATAAAAATACAATTTATTAAATAATATTATTACAAAAAAATAATAAAAAAAAA 5996

Qy      1014 GTCATAATTGCTGAAGAATTAAAGA-----AGTGGAACCTTAAGATTCATGAAGGACCTA 1068
             | ||||    || ||  |||| |     | |  |  |  |  ||| ||  |     |
Db      5995 AAAAAAATTTAAAAAAAAATAAAAAAAAATAATAAACACACATAATTAATTTATACAATT 5936

Qy      1069 TTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAG 1128
             |||||     ||  |    ||| ||| ||||  | |  ||| |  |    | |||||  ||
Db      5935 TTTTTATTTTATTTATATTTTAAAATTAAAATATTCAATTATTCAATTAAATTTTTTTTT 5876

Qy      1129 TAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGA 1188
             ||  |   |  | ||  |   | | ||||      |||  | |    |||  |||   |
Db      5875 TATTTTTTTTATTTTTTTTATTTTTTTTATAAACATTTTTTTCAAATTTTTAAAATAAAA 5816
```

Qy 1189 GAAAACAGTCTTAACAAAAAAAA 1211
| | | || |||||||
Db 5815 AACATAAAAAAAAAAAAAAAAAAA 5793

RESULT 8
ABZ10188
ID ABZ10188 standard; DNA; 1501 BP.
XX
AC ABZ10188;
XX
DT 16-JAN-2003 (first entry)
XX
DE Haematopoietic cell proliferation disorder related DNA sequence #328.
XX
KW Human; haematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
PD 03-OCT-2002.
XX
PF 26-MAR-2002; 2002WO-EP003401.
XX
PR 26-MAR-2001; 2001US-0278333P.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
PI Schwope I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
PT Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent that
PT distinguishes between methylated and non-methylated CpG dinucleotides.
XX
PS Claim 28; SEQ ID NO 328; 117pp; English.
XX
CC The present invention describes a method for detecting and
CC differentiating between haematopoietic cell proliferative disorders
CC associated with at least 1 gene and/or their regulatory regions in a
CC subject. The method comprises contacting a target nucleic acid in a
CC biological sample obtained from the subject with at least 1 reagent,
CC which distinguishes between methylated and non-methylated CpG
CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related DNA
CC sequences. The nucleotide sequences from the present invention can also
CC be used for detecting a predisposition to, differentiation between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables a
CC highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients
XX
SQ Sequence 1501 BP; 604 A; 0 C; 130 G; 767 T; 0 U; 0 Other;

Query Match 4.4%; Score 53.2; DB 8; Length 1501;
Best Local Similarity 46.1%; Pred. No. 0.52;
Matches 259; Conservative 0; Mismatches 293; Indels 10; Gaps 2;

```
Qy       663 TTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGA 722
             ||  |  |||| || |  |||||  | |  || ||    | |  |||   |||||  |
Db       900 TTAATGATTTTTAATTTTAAATTTAAATTAATTTTGATGTTTGTTTAATTATTTGTTGAT 959

Qy       723 ATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTT 782
              | ||  ||   ||  |     ||   |||   ||     |  |  | ||  | |   ||
Db       960 TTGTTTTTTAAATTTAATAATTTAAATTTAAATAATATTGAAATGAAAATTTTGAAATTT 1019

Qy       783 TCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAA 842
                |  | ||    ||| || | |  |  | || ||     | |  || ||||   |
Db      1020 AAAATTTGAAATGTAATTTTTTTAAATTTTAATTTTATT----TATAAAATAAAAAAATA 1075

Qy       843 ATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGA 902
             || | |||      |  ||    ||    | ||||  || |  ||| |  |  ||  | |
Db      1076 ATAATATTTATTTTATAAAATTATTATAAAATTTTAAAAAAATAAAAATATTTAAAAAAA 1135

Qy       903 AACTACGCTTTCCACGACATATTCTCTCTACTT------ATTATGGGTGGTATGTAATAT 956
             || |    |||  |  | | | |   |  ||     ||  |  | |||    | |  |  | |
Db      1136 AAATTTAATTTAAATAAAAAAATAAATGATGTTAATTTAATTTTTGTTTTAAAAAAATGA 1195

Qy       957 TAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTC 1016
             || |   | || |||| | || |   |  ||| |   |  ||     |||||   ||
Db      1196 TATAAAAATAAAAAATTTTTTTTAAAATAAAATTAATTAAAATTTGTTATATTTTTTTTT 1255

Qy      1017 ATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAA 1076
             ||  |   | ||  |||| | || |   |  ||| |   |  ||     |||||   ||
Db      1256 ATTTTATATAAAATATTATAAAAATTATATAATAAAAAATAAAAATTTAATATTTAAAAA 1315

Qy      1077 AAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTA 1136
             | ||   |   || ||      |||||||  ||| |    ||   |||  |  |  | ||
Db      1316 ATATTATAAAATTAAATTATTTTAAATTTTATTTTATATTATTAATTAAAATAATAAATA 1375

Qy      1137 ACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAG 1196
             |   || ||   ||| | || |      ||||  |||   || ||||      |||
Db      1376 ATTTTTTTATATAATTTTTTTTTTAAATTTTAAATATATTTTATTAAAAAATAAAATTTT 1435

Qy      1197 TCTTAACAAAAAAAAAAAAAAA 1218
               ||||   ||||  |||| |
Db      1436 ATTTAATGTTAAAATTAAAATA 1457
```

RESULT 9
ADE84162
ID ADE84162 standard; DNA; 1501 BP.
XX
AC ADE84162;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human lymphoid cell proliferative disorder gene derived DNA #98.
XX
KW ds; lymphoid cell proliferative disorder; methylation;
KW methylated CpG dinucleotide; single nucleotide polymorphism; SNP;
KW diffuse large B-cell lymphoma; mantle cell lymphoma;
KW chronic lymphocytic leukemia; small lymphocytic lymphoma;
KW follicular lymphoma; diagnosis; prognosis.
XX
OS Homo sapiens.
XX
PN WO2003044226-A2.
XX
PD 30-MAY-2003.
XX
PF 25-NOV-2002; 2002WO-EP013265.
XX
PR 23-NOV-2001; 2001DE-01057491.
PR 28-DEC-2001; 2001DE-01064501.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Burger M, Caldwell C, Genc B, Becker E, Maier S, Nimmrich I;
XX
DR WPI; 2003-457621/43.
XX
PT Detecting and differentiating between lymphoid cell proliferative
PT disorders comprises contacting a target nucleic acid with at least one
PT reagent that distinguishes between methylated and non-methylated CpG
PT dinucleotides.
XX
PS Claim 26; SEQ ID NO 158; 448pp; English.
XX
CC The invention relates to a method of detecting and differentiating
CC between lymphoid cell proliferative disorders associated with at least
CC one gene and/or their regulatory regions in a subject by contacting a
CC target nucleic acid in a biological sample obtained from the subject with
CC at least one reagent or series of reagents that distinguish between
CC methylated and non-methylated CpG dinucleotides within the target nucleic
CC acid. The genes and/or their regulatory regions are preferably selected
CC from MDR1, CSNK2B, EGR4, AR, CDK4, RB2, CDC25A, GPIb beta, MYOD1, CDH3,
CC MYCL1, ELK1, ABL1, APC, BCL2, CDH1, CDKN1A, CDKN1B, CDKN2a, CDKN2B, FOS,
CC GSTP1, HIC-1, MGMT, MLH1, MOS, MYC, PTEN, RBL2, TGFBR2, TP73, CDKNIC,
CC GSK3beta, ESR1, APAF1, BAK1, BAX or HOXA5. Oligomers, peptide nucleic
CC acid (PNA)-oligomers and/or isolated nucleic acids based on the sequences
CC of the genes are useful for detecting the methylation state of all the
CC CpG dinucleotides within one or more the sequences, or their complements,
CC for determining the cytosine methylation state and or single nucleotide
CC polymorphisms (SNPs), and for differentiating at least two of the medical
CC conditions such as diffuse large B-cell lymphoma, mantle cell lymphoma,
CC chronic lymphocytic leukemia, small lymphocytic lymphoma and follicular
CC lymphoma. They are also useful for detecting of a predisposition to,
CC differentiation between subclasses, diagnosis, prognosis, treating and/or
CC monitoring of lymphoid cell proliferative disorder. This sequence
CC represents a nucleic acid of a pretreated genomic DNA derived from the
CC above mentioned genes.
XX
SQ Sequence 1501 BP; 604 A; 0 C; 130 G; 767 T; 0 U; 0 Other;

Query Match 4.4%; Score 53.2; DB 10; Length 1501;
Best Local Similarity 46.1%; Pred. No. 0.52;
Matches 259; Conservative 0; Mismatches 293; Indels 10; Gaps 2;


```
Qy       663 TTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGA 722
             ||  |  |||| || |  |||||  | |  || ||    | |  |||   |||||  |
Db       900 TTAATGATTTTTAATTTTAAATTTAAATTAATTTTGATGTTTGTTTAATTATTTGTTGAT 959

Qy       723 ATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTT 782
              | ||  ||   ||  |     ||   |||   ||     |  |  | ||  | |   ||
Db       960 TTGTTTTTTAAATTTAATAATTTAAATTTAAATAATATTGAAATGAAAATTTTGAAATTT 1019

Qy       783 TCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAA 842
                |  | ||    ||| || | |  |  | || ||     | |  || ||||   |
Db      1020 AAAATTTGAAATGTAATTTTTTTAAATTTTAATTTTATT----TATAAAATAAAAAAATA 1075

Qy       843 ATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGA 902
             || | |||      |  ||    ||    | ||||  || |  ||| |  |  ||  | |
Db      1076 ATAATATTTATTTTATAAAATTATTATAAAATTTTAAAAAAATAAAAATATTTAAAAAAA 1135

Qy       903 AACTACGCTTTCCACGACATATTCTCTCTACTT------ATTATGGGTGGTATGTAATAT 956
             || |    |||  |  | | | |   |  ||     ||  |  | |||    | |  |  | |
Db      1136 AAATTTAATTTAAATAAAAAAATAAATGATGTTAATTTAATTTTTGTTTTAAAAAAATGA 1195

Qy       957 TAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTC 1016
             || |   | || |||| | || |   |  ||| |   |  ||     |||||   ||
Db      1196 TATAAAAATAAAAAATTTTTTTTAAAATAAAATTAATTAAAATTTGTTATATTTTTTTTT 1255

Qy      1017 ATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAA 1076
             ||  |   | ||  |||| | || |   |  ||| |   |  ||     |||||   ||
Db      1256 ATTTTATATAAAATATTATAAAAATTATATAATAAAAAATAAAAATTTAATATTTAAAAA 1315

Qy      1077 AAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTA 1136
             | ||   |   || ||      |||||||  ||| |    ||   |||  |  |  | ||
Db      1316 ATATTATAAAATTAAATTATTTTAAATTTTATTTTATATTATTAATTAAAATAATAAATA 1375

Qy      1137 ACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAG 1196
             |   || ||   ||| | || |      ||||  |||   || ||||      |||
```

```
Db       1376 ATTTTTTTATATAATTTTTTTTTAAATTTTAAATATATTTTATTAAAAAATAAAATTTT 1435

Qy       1197 TCTTAACAAAAAAAAAAAAAA 1218

Db       1436 ATTTAATGTTAAAATTAAAATA 1457


RESULT 10
AAA02542/c
ID   AAA02542 standard; cDNA; 1199 BP.
XX
AC   AAA02542;
XX
DT   19-MAY-2000  (first entry)
XX
DE   Human colon cancer cell line polynucleotide sequence SEQ ID NO:2533.
XX
KW   Human; colon cancer; tumour; diagnosis; gene expression product; probe;
KW   detection; cancerous state; metastasis; identification; breast cancer;
KW   oestrogen receptor-positive breast cancer; therapy;
KW   oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9958675-A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999;   99WO-US010602.
XX
PR   14-MAY-1998;   98US-0085426P.
PR   15-MAY-1998;   98US-0085537P.
PR   15-MAY-1998;   98US-0085696P.
PR   21-OCT-1998;   98US-0105234P.
PR   27-OCT-1998;   98US-0105877P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Giese K,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A;
PI   Lamson G,  Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones LW,  Stache-Crain B;
XX
DR   WPI; 2000-126369/11.
XX
PT   Polynucleotide library used to determine cancerous states of mammalian
PT   cells.
XX
PS   Claim 1; Page 1023-1024; 1097pp; English.
XX
CC   AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC   libraries constructed from human colon cancer cell lines. The present
CC   invention also describes a method of detecting differentially expressed
CC   genes correlated with a cancerous state of a mammalian cell, comprising
CC   detecting at least one differentially expressed gene product in a test
CC   sample derived from a cell suspected of being cancerous, where detection
CC   of the differentially expressed gene product is correlated with a
CC   cancerous state of the cell from which the test sample was derived. The
CC   polynucleotides sequences can be used in a method for detecting
CC   differentially expressed genes correlated with a cancerous state of a
CC   mammalian cell. The polynucleotides can also be used as probes for
CC   detecting and mapping related genes. They can be used in diagnosis and
CC   prognosis of diseases and disorders (e.g. identification of pre-
CC   metastatic or metastatic cancerous states, stages of cancer, or
CC   responsiveness of cancer to therapy). This is particularly for breast
CC   cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC   negative breast cancer, lung cancer, and colon cancer
XX
SQ   Sequence 1199 BP; 430 A; 98 C; 63 G; 370 T; 0 U; 238 Other;

  Query Match             4.3%;  Score 52.6;  DB 3;  Length 1199;
  Best Local Similarity   37.7%;  Pred. No. 0.67;
  Matches  268;  Conservative    0;  Mismatches  437;  Indels    6;  Gaps    3;

Qy        486 TCTATGAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTCCTCTT 545

Db       1180 TNTATANNNTAGAGANANANTTANTNATAATATATATTTNNNNAAATNNAGNATAATGTT 1121

Qy        546 TGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGTTCACA 605

Db       1120 TATAATTGTATNNAATAGTATTGATGAANTAGATAGATANATTNATAT-ATGTNAANANA 1062

Qy        606 AACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCCCCTTC 665

Db       1061 NATTATATATGAAANANGTAAGATNTATATATACATATNTNNTTATNTTNTAATNNAATA 1002

Qy        666 TTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGAATC 725

Db       1001 TNANTNNGTTCNTATNATCTATATATANNAANAGCTTATNAATATATATANTNAGTAANA 942

Qy        726 TTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCT 785

Db        941 TTAAATTTATT-ATNTANNTNTATTTATANTAGATATNTATTATNTNAATAANTNNATAT 883

Qy        786 GAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATG 845

Db        882 NATTTTAANNTAGTTATNNGTAGANNATANTATAAGTTTANATNTAAATTAAATNANNTG 823

Qy        846 AAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAAC 905

Db        822 TNAATAATANGATTTNATANNTNTCTATGAAAANTNNATGAATGATNTAAATAAATGTAG 763

Qy        906 TACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAA 965

Db        762 TTCTATTANTAGTNGATATTTNTTTTAANTATTTNTNGATATAAGTANTTNTNCNTATAT 703

Qy        966 CATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCT 1025

Db        702 TTNTNNANAATNTNATANATNNTTTTATNNTNGATAANTATATATTATAAATTAATTNNN 643

Qy       1026 GA----AGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATC 1081

Db        642 TATNTNATTATNAATCAAATNTTNTNTGTATTATNTNGNTATNAAANTATANGAGNTNTG 583

Qy       1082 TCACCTTTGAAAATAAATAAATTTATTTAAATGCATATTTTATTAGTAATACTAACAAT 1141

Db        582 TTTNTTTAAGTTATTAATAATNATNTATTGTAANGATATNTATATTANNATTTNNANNAT 523

Qy       1142 TGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAA 1192

Db        522 GANTANATANTTTATATNTAGTAAGTAATAATTTTCGATTNTTNTNATATA 472


RESULT 11
ABA92787_0
WP Sequence split into 7 fragments  LOCUS ABA92787 Accession Aba92787
WP      Fragment Name      Begin     End
WP      ABA92787_0             1     110000
WP      ABA92787_1        100001     210000
WP      ABA92787_2        200001     310000
WP      ABA92787_3        300001     410000
WP      ABA92787_4        400001     510000
WP      ABA92787_5        500001     610000
WP      ABA92787_6        600001     640681
ID   ABA92787 standard; DNA; 640681 BP.
XX
AC   ABA92787;
XX
DT   27-MAR-2002  (first entry)
XX
DE   Buchnera sp. genomic DNA SEQ ID NO:1.
XX
KW   Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
KW   circular; ds.
XX
```

OS Buchnera sp.
XX
PN JP2001292771-A.
XX
PD 23-OCT-2001.
XX
PF 07-APR-2000; 2000JP-00107160.
XX
PR 07-APR-2000; 2000JP-00107160.
XX
PA (RIKA) RIKAGAKU KENKYUSHO.
XX
DR WPI; 2002-126043/17.
XX
PT A genomic DNA of cockroach-symbiotic bacterium.
XX
PS Claim 1; Page 16-230; 237pp; Japanese.
XX
CC The present invention describes a gene (I) derived from Buchnera sp.
CC containing the DNA (a) or (b), (a) has a fully defined base pair sequence
CC selected from a table of sequences found in the Buchnera sp. genomic DNA
CC of ABA92787 given in the specification or is a DNA selected from
CC complementary DNA sequences, and (b) is a DNA which hybridises with the
CC DNA (a) and encodes a protein. Also described are: (1) a recombinant
CC vector (II) containing (I); (2) a transformant (III) containing (II); (3)
CC a genomic DNA of Buchnera sp. containing the sequence given in ABA92787;
CC (4) a plasmid derived from Buchnera sp. containing DNA (c) or (d), (c) is
CC a DNA containing a fully defined sequence given in ABA92788 or ABB92789
CC and (d) is a plasmid which hybridises with a DNA; and (5) a method for
CC the preparation of a protein in which (III) is cultured and the
CC expression protein of the objective protein is collected from the
CC resultant culture. The DNA is useful for developing agricultural
CC chemicals for exterminating cockroaches. The present sequence represents
CC the specifically claimed Buchnera sp. genomic DNA sequence, from the
CC present invention
XX
SQ Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 U; 0 Other;

Query Match 4.3%; Score 52; DB 6; Length 110000;
Best Local Similarity 54.8%; Pred. No. 1.4;
Matches 103; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy       1031 ATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTG 1090
             | |||| || |    |   ||  ||  || |  ||  ||||||| ||||||||     |
Db      95585 AGTAAAAAAATATTAATATACTATCAATAATTGAAATATTTTTTAAAAATCTTTTTCTAT 95644

Qy       1091 AAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAA 1150
              || ||| | |||| || || ||     |  ||||  | | |  ||  ||| ||  |
Db      95645 TTAAAAAACATATTTTTTATATATCTGACTAGTATTTTTTAAAAAAATCATTTTAATATT 95704

Qy       1151 TGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAA 1210
              | | | ||  |     | | || ||| | || |     |  |  |||| |  | |||| ||
Db      95705 TTTTTGATATTAAAATGATTTTTTATTCTTAATATATATTATTCAGTTTATATAAAATAA 95764

Qy       1211 AAAAAAAA 1218
              |||| | |
Db      95765 AAAATATA 95772
```

RESULT 12
ABX46053/c
ID ABX46053 standard; cDNA; 424 BP.
XX
AC ABX46053;
XX
DT 21-FEB-2003 (first entry)
XX
DE Bovine EST associated with lactation/muscle/fat deposition #11218.
XX
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW muscle deposition; fat deposition; genome mapping; gene identification;
KW gene analysis; cattle breeding.
XX

OS Bos Taurus.
XX
PN US2002137139-A1.
XX
PD 26-SEP-2002.
XX
PF 24-SEP-2001; 2001US-00960352.
XX
PR 12-JAN-1999; 99US-0115707P.
PR 11-JAN-2000; 2000US-00480902.
XX
PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR WPI; 2003-110599/10.
XX
PT New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and analysis,
PT cattle breeding, or for genetically improving cattle.
XX
PS Claim 2; SEQ ID NO 11218; 245pp; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived from
CC cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC appearing as ABX34836-ABX49947, or complements of them. Also included are
CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC acid linked to a promoter and a 3' non- translated sequence that
CC functions in the cell to cause termination of transcription and addition
CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC (2) determining a level or pattern of a molecule in a bovine cell or
CC tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC of the 15112 nucleic acid sequences or its complement or fragment) with a
CC complementary nucleic acid molecule obtained from the bovine cell or
CC tissue, where hybridisation between the marker nucleic acid and the
CC complementary nucleic acid permits the detection of the molecule; and (b)
CC detecting the level or pattern of the complementary nucleic acid, where
CC the detection of the complementary nucleic acid is predictive of the
CC level or pattern of the molecule. The LMFD nucleic acid is used for
CC determining a level or pattern of a molecule in a bovine cell or tissue.
CC It is useful for genome mapping, gene identification and analysis, cattle
CC breeding, preparation of constructs for use in cattle gene expression, or
CC for genetically improving cattle. The present sequence is one of the
CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC present sequence was not shown in the specification but was obtained in
CC electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
SQ Sequence 424 BP; 228 A; 7 C; 32 G; 157 T; 0 U; 0 Other;

Query Match 4.2%; Score 51.8; DB 8; Length 424;
Best Local Similarity 63.0%; Pred. No. 0.85;
Matches 80; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy       1093 AATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATG 1152
              || |  | ||||||||||||  |  | |||||  ||  |  |  | |   || |      |
Db        194 AAAATTTTAATTTATTTTATTTTTAAATTTTTTTTATTTTTTTTATTTTTTTTTTTTTT 135

Qy       1153 TGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAA 1212
              | || |  ||| |||| || ||  |||||||  | | |||| |     || |||||||||
Db        134 TATTTTTTTTTTTTTATTTTTTTATTTTTAATTTAAAAAAAAAAAAAAAAAAAAAAAAA 75

Qy       1213 AAAAAAA 1219
              |||||||
Db          74 AAAAAAA 68
```

RESULT 13

```
ABL34044/c
ID   ABL34044 standard; DNA; 7319 BP.
XX
AC   ABL34044;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2017.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 2017; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 7319 BP; 2188 A; 263 C; 1633 G; 3235 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 51.8;  DB 6;  Length 7319;
  Best Local Similarity   48.7%;  Pred. No. 1.2;
  Matches  170;  Conservative    0;  Mismatches  177;  Indels    2;  Gaps    1;

Qy          3 TAAGCAAACTAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAATT 62
              ||| | |  |||||  | | || || ||||       || | || ||  ||  | ||| |
Db       1875 TAAACCACTTAAATCCATATAAAAATCATTTAAAATTTTAACAAAACACAAACAAAAAAT 1816

Qy         63 GTTCTTGAAAATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTGAA 122
                |||| ||||||  ||  | | ||  | | |    | ||    || || ||  ||||  ||
Db       1815 TATCTTTAAAATATACACAACCTAATAATCATTAAAAACAAAATCTAAAACTATACACAA 1756

Qy        123 TATGACTGCACTGAATATGCTAAATCAAA--TAAGCATCCTGGCGGTCTTAATTTCCTCA 180
               ||       | |  | |     ||| || |  ||| ||||       |   ||||||  |
Db       1755 TAATTACCCCCAAATTTAATAAAACCATAATTAACCATCAATTATTTAAAAATTTCACCG 1696

Qy        181 ATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGT 240
               |  || |||  | | ||  |     |  | |  | | |  | | |||  || |  || |
Db       1695 CATAATTCATTTCTAAAAAAAATATATAAATTATAAAATCAATATCACAACAATAAAATT 1636

Qy        241 AGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAAT 300
```

```
                    ||||    |   |   ||||  | |||    ||||        | |   | |
Db       1635 TAAAAAATAAAAAAAAATCCTTAATCCCACAAACTAATACAAATATCTTTTAAATATATT 1576

Qy        301 CTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGA 349
              |   | | | | | |  | | |||| || ||    ||    | ||||  |
Db       1575 CACTATATATAATTTTTACAATAAAAAACAAAAAAAACTTATAATTCTA 1527


RESULT 14
ABL32101/c
ID   ABL32101 standard; DNA; 5728 BP.
XX
AC   ABL32101;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 74.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW   ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP007537.
XX
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful for
PT   diagnosis and treatment of diseases associated with abnormal cytosine
PT   methylation.
XX
PS   Claim 1; SEQ ID NO 74; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention
XX
SQ   Sequence 5728 BP; 1931 A; 49 C; 885 G; 2863 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 51.4;  DB 6;  Length 5728;
  Best Local Similarity   54.7%;  Pred. No. 1.4;
  Matches  123;  Conservative    0;  Mismatches  101;  Indels    1;  Gaps    1;

Qy        989 TTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCT 1048
              | |  ||  |  | | |||   | | ||| ||      || ||  ||| ||        |
Db       2910 TCTAAACGAAAAAACGAAACTCCATCTCAAAAAAAAAAAAAAAAAAAAAAAAATTCTTT 2851

Qy       1049 TAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATT 1108
              |||  ||||| ||      | ||       ||| |||  |  || || ||||| ||||  |||
Db       2850 TAAATTTCATCAATTTTATTTTAAAACTAAAACTCTTCATTAAATATAAAAAAATCAATT 2791
```

```
Qy       1109 TTAAATGCATATTTTATTAGTA-ATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTT 1167
              ||||   ||||  || |  | | | | | |||| |   |||| | ||    |||
Db       2790 CAAAATTTCTATTCAATCATAATAAATTTAAAATTTTTTTAAATTTCTTTATATTTAACA 2731

Qy       1168 ACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAA 1212
              |  |||||| || |||| | | || | |  || || |||  || |
Db       2730 ATATATTACATTATAATATTATAATAAAAACTAAATAAACTAATA 2686


RESULT 15
AAS46709/c
ID   AAS46709 standard; DNA; 5750 BP.
XX
AC   AAS46709;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #432.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic; cancer;
KW   tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP002955.
XX
PR   15-MAR-2000; 2000DE-01013847.
PR   06-APR-2000; 2000DE-01019058.
PR   07-APR-2000; 2000DE-01019173.
PR   30-JUN-2000; 2000DE-01032529.
PR   01-SEP-2000; 2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumor suppressor
PT   genes and oncogenes, useful in designing primers and probes for analyzing
PT   diseases associated with cytosine methylation state e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 432; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and oncogenes
CC   having a sequence taken from 536 (actually 533 since numbers 408, 458 and
CC   500 are missing from the sequence listing) sequences (Ss) and sequences
CC   complementary to (Ss). The nucleic acid may be a peptide nucleic acid-
CC   oligomer (PNA) of at least 9 nucleotides and may form part of a set of
CC   probes for detecting the cytosine methylation state and/or single
CC   nucleotide polymorphisms and also to be used in an array for analysing
CC   diseases associated with CpG dinucleotides e.g. cancers and tumours. The
CC   probes can also be used in a method for ascertaining genetic and/or
CC   epigenetic parameters for the diagnosis and/or therapy of existing
CC   diseases or the predisposition to specific diseases, by analysing
CC   cytosine methylations. The parameters may be compared to another set of
CC   genetic and/or epigenetic parameters, the differences serving as basis
CC   for diagnosis and/or prognosis events which are disadvantageous to
CC   patients. The present sequence is one of the 533 genomic sequences
CC   derived from tumour suppressor genes and oncogenes. Sequences with even
CC   numbered Seq ID numbers are the complementary sequence of the
CC   corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
CC   535, except for those whose partner sequence is missing). Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 5750 BP; 1736 A; 103 C; 1392 G; 2519 T; 0 U; 0 Other;

  Query Match             4.2%;  Score 51.2;  DB 4;  Length 5750;
  Best Local Similarity   45.7%;  Pred. No. 1.5;
  Matches  179;  Conservative    0;  Mismatches  213;  Indels    0;  Gaps    0;

Qy        827 AAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCA 886
              ||||  | ||||| |||  ||   |   | |      |       || |||| ||||  |
Db       4820 AAATATTTAAAATAAAAATAACGACACTAAATCTTAAAAAAATAAATCTTTCAAACAAAA 4761

Qy        887 AATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGG 946
              || |  |  || ||  || ||  |      |  | ||   |||| | |
Db       4760 AACAAATATAAACAATAAATATCCCACGAAAAAAATTCCCTCTAAAATACGCCCTAAAAA 4701

Qy        947 TATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAA 1006
              | | || | ||   |  | |  |||| ||   | || || || | |  | ||  | |
Db       4700 AACGAAAAAATATTTTATAAAAAACTTATTTTAAAAACTTTCTCTTTCTAAATATCTCCT 4641

Qy       1007 AGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACC 1066
                || |  ||  | ||   || |      |  ||||| || ||   | |
Db       4640 TTATCCTATTTAAAATTATTAAACCTTCATACTATTTAACCTCAAAATATTTAACTACTT 4581

Qy       1067 TATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATT 1126
              |  | |  |||||   ||||  | |  |   | ||    ||   |  |  | |||   ||
Db       4580 AAAATATTCAAAATTAAACCTATAATCAATCAAAATAAAATAAAATCTATTTCTTAATTT 4521

Qy       1127 AGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCT 1186
              | | ||       | || |     || |  | |  ||  ||   || |||   |  |||||
Db       4520 ATTTATTTATTTATTTTTTTTTATTTTATTTATTTATTTATTTTTTTAAAATAAAATCT 4461

Qy       1187 GAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1218
                |  | |  ||| |||| | || || |
Db       4460 CGCCCTATCGCCCAAACTAAAATACAATAACA 4429


Search completed: January 19, 2005, 14:11:06
Job time : 685 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: January 18, 2005, 21:56:08 ; Search time 5526 Seconds
(without alignments)
10431.803 Million cell updates/sec

Title: US-10-070-666A-1
Perfect score: 1219
Sequence: 1 agtaagcaaactaaatttaa.....taacaaaaaaaaaaaaaaaa 1219

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1219	100.0	1219	6	AX098429	AX098429 Sequence
2		614.4	50.4	2492	6	AX098431	AX098431 Sequence
c 3		64.2	5.3	266544	3	AC116956	AC116956 Dictyoste
4		64	5.3	2000	6	AX655393	AX655393 Sequence
c 5		62.4	5.1	252632	3	AE014818	AE014818 Plasmodiu
6		59	4.8	8056	6	AX599046	AX599046 Sequence
7		58.4	4.8	1141	6	AX083744	AX083744 Sequence
c 8		58	4.8	6103	6	AX346592	AX346592 Sequence
9		58	4.8	253001	3	AE014834	AE014834 Plasmodiu
c	10	57.8	4.7	3026	3	AF324424	AF324424 Ichthyoph
c	11	57.4	4.7	8056	6	AX598900	AX598900 Sequence
c	12	57	4.7	331039	3	AC116979	AC116979 Dictyoste
c	13	56.8	4.7	7079	3	AF362375	AF362375 Dictyoste
	14	56.8	4.7	159835	2	CR387984	CR387984 Danio rer
c	15	56.4	4.6	76568	3	MBREV	AF538053 Monosiga
	16	56.2	4.6	111882	3	AC115612	AC115612 Dictyoste
c	17	55.6	4.6	88549	3	AC116924	AC116924 Dictyoste
	18	55.4	4.5	110000	2	PFMAL13_14	Continuation (15 o
	19	55.2	4.5	250029	3	AE014838	AE014838 Plasmodiu
	20	55.2	4.5	250053	3	AE014825	AE014825 Plasmodiu
c	21	54.8	4.5	3606	3	DDI1093A	M19469 Dictyosteli
	22	54.8	4.5	125623	3	AC115599	AC115599 Dictyoste
	23	54.8	4.5	229947	2	CR394564	CR394564 Danio rer
c	24	54.6	4.5	8056	6	AX599046	AX599046 Sequence
c	25	54.6	4.5	110000	1	AP006628_6	Continuation (7 of
c	26	54.4	4.5	1141	6	AX083744	AX083744 Sequence
c	27	54.4	4.5	134768	5	BX537293	BX537293 Zebrafish
	28	54.2	4.4	348174	3	CR382399	CR382399 Plasmodiu
	29	54	4.4	183285	2	BX899180	BX899180 Danio rer
	30	54	4.4	281723	3	PFA929359	AL929359 Plasmodiu
	31	53.4	4.4	110000	2	PFMAL13_06	Continuation (7 of
c	32	53.4	4.4	197225	9	AC093835	AC093835 Homo sapi
c	33	53.4	4.4	341050	3	PFA929357	AL929357 Plasmodiu
	34	53.2	4.4	1501	6	AX598988	AX598988 Sequence
	35	53.2	4.4	1501	6	AX767510	AX767510 Sequence
c	36	53.2	4.4	6099	3	AF482382	AF482382 Dictyoste
c	37	53.2	4.4	180915	8	GTAJ10592	AJ010592 Guillardi
	38	53.2	4.4	190015	2	AC140056	AC140056 Rattus no
	39	53	4.3	1192	9	HSA323759	AJ323759 Homo sapi
	40	53	4.3	104992	2	AC005504	AC005504 Plasmodiu
	41	53	4.3	169546	2	AC004157	AC004157 Plasmodiu
c	42	53	4.3	250421	3	AE014849	AE014849 Plasmodiu
	43	52.8	4.3	156060	2	AC004153	AC004153 Plasmodiu
	44	52.8	4.3	250707	3	AE014848	AE014848 Plasmodiu
c	45	52.6	4.3	1199	6	BD221420	BD221420 Human gen

ALIGNMENTS

RESULT 1
AX098429
LOCUS AX098429 1219 bp DNA linear PAT 02-APR-2001
DEFINITION Sequence 1 from Patent WO0120000.
ACCESSION AX098429
VERSION AX098429.1 GI:13537718
KEYWORDS .
SOURCE Tetrahymena thermophila
ORGANISM Tetrahymena thermophila
Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE 1
AUTHORS Ruesing,M., Kiy,T. and Dominitzki,A.
TITLE Nucleic acid which is obtained from tetrahymena and which codes for a delta-6-desaturase, the production thereof and use
JOURNAL Patent: WO 0120000-A 1 22-MAR-2001;
Axiva GmbH (DE)
FEATURES Location/Qualifiers
source 1. .1219
/organism="Tetrahymena thermophila"
/mol_type="unassigned DNA"
/db_xref="taxon:5911"
ORIGIN

Query Match 100.0%; Score 1219; DB 6; Length 1219;
Best Local Similarity 100.0%; Pred. No. 4.2e-202;
Matches 1219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGTAAGCAAACTAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTAAGCAAACTAAATTTAAAAAACAAGCATTATGGGAGTTGATAAGACTTAAGAAGAAA 60

Qy         61 TTGTTCTTGAAAATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTGTTCTTGAAAATAAACCCGAACTTCTCAACGAATACAAATTTATTTACAAGGATACTG 120

Qy        121 AATATGACTGCACTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AATATGACTGCACTGAATATGCTAAATCAAATAAGCATCCTGGCGGTCTTAATTTCCTCA 180

Qy        181 ATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 ATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCAGAACACTCCATTCTAAGT 240

Qy        241 AGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAAAATAAGAGGAGACTGAAT 300

Qy        301 CTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGCATTTATTCGAACCAAACTGGC 360

Qy        361 CTATCGAAATTGGTTTATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTATCGAAATTGGTTTATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTC 420

Qy        421 AAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTG 480

Qy        481 GTCACTCTATGAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTCACTCTATGAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTC 540

Qy        541 CTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGT 600

Qy        601 TCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCC 660

Qy        661 CCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAG 720

Qy        721 GAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTC 780

Qy        781 TTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATG 840

Qy        841 AAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCA 900

Qy        901 GAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGA 960

Qy        961 CTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAA 1020

Qy       1021 TTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAAT 1080

Qy       1081 CTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAA 1140

Qy       1141 TTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTT 1200

Qy       1201 AACAAAAAAAAAAAAAAAA 1219
              ||||||||||||||||||||
Db       1201 AACAAAAAAAAAAAAAAAA 1219
```

RESULT 2
AX098431

```
LOCUS       AX098431               2492 bp    DNA     linear   PAT 02-APR-2001
DEFINITION  Sequence 3 from Patent WO0120000.
ACCESSION   AX098431
VERSION     AX098431.1  GI:13537719
KEYWORDS    .
SOURCE      Tetrahymena thermophila
  ORGANISM  Tetrahymena thermophila
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Tetrahymenina; Tetrahymena.
REFERENCE   1
  AUTHORS   Ruesing,M., Kiy,T. and Dominitzki,A.
  TITLE     Nucleic acid which is obtained from tetrahymena and which codes for
            a delta-6-desaturase, the production thereof and use
  JOURNAL   Patent: WO 0120000-A 3 22-MAR-2001;
            Axiva GmbH (DE)
FEATURES             Location/Qualifiers
     source          1. .2492
                     /organism="Tetrahymena thermophila"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5911"
ORIGIN

  Query Match             50.4%;  Score 614.4;  DB 6;  Length 2492;
  Best Local Similarity   99.8%;  Pred. No. 3.6e-97;
  Matches  615;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy         589 AACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAAT 648
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1652 AGCATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAAT 1711

Qy         649 TGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATT 708
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1712 TGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATT 1771

Qy         709 ATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCT 768
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1772 ATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCT 1831

Qy         769 ATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAA 828
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1832 ATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAA 1891

Qy         829 ATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAA 888
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1892 ATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAA 1951

Qy         889 TAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTA 948
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1952 TAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTA 2011

Qy         949 TGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAG 1008
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2012 TGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAG 2071

Qy        1009 CTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTA 1068
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2072 CTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTA 2131

Qy        1069 TTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAG 1128
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2132 TTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATATTTTATTAG 2191

Qy        1129 TAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGA 1188
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2192 TAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGA 2251

Qy        1189 GAAAACAGTCTTAACA 1204
               ||||||||||||||||
Db        2252 GAAAACAGTCTTAACA 2267
```

RESULT 3
AC116956/c

```
LOCUS       AC116956              266544 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map 1418423-1684967 strain
            AX4, complete sequence.
ACCESSION   AC116956 AC116980 AC115678
VERSION     AC116956.2  GI:28828496
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 266544)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 266544)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 4, 2003 this sequence version replaced
            gi:20042935, gi:19682974, gi:19919988.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .266544
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="1418423-1684967"
     CDS              join(<842. .1240,1322. .1916)
                      /note="GeneID exon scores (in order of location ranges):
                      29.76, 65.85 - GSCJ_ID dd_02843"
                      /codon_start=2
                      /product="similar to Dictyostelium discoideum (Slime
                      mold). Hypothetical 98.7 kDa protein (Fragment)"
                      /protein_id="AAO51167.1"
                      /db_xref="GI:28828559"
                      /translation="VAISNFNINNNNNKNNNDNKEIRFPTDIIIDDSTEINFNNHEDNN
                      NNNNNNNNNNAYNEDSGNSVQLGDIECVVCLDFLDSYLPTLVKIISEYGVIESCSKI
                      CGLLNKTDEIDICTGLCDLVGVDTFWKIFVTNDINPFYACELITACSTPKNPAADFIS
                      LGVSPAIGQSGSKFLIDLTFQVVNTTGAGQFAFIVFYTQTQSKFINYQLFEGFSPGNY
                      SVKYDFETTNNSTFINGQYPVTVYMCNGECGTMYSTLLNQTTSYFTISNPTPTPTPTP
                      SNSTTPTPTPTNSTPTPTSTSTPTSTPTSTPTPTPTSSSSTKTHSSHYKNKINK"
     CDS              complement(join(2119. .2662,2752. .2943,3024. .3055))
                      /note="GeneID exon scores (in order of location ranges):
                      32.18, 11.63, -1.19 - GSCJ_ID dd_02844"
                      /codon_start=1
                      /product="similar to Mus musculus (Mouse). 16 days neonate
                      thymus cDNA, RIKEN full-length enriched library,
                      clone:A130067E11 product:cyclin C, full insert sequence"
                      /protein_id="AAO51105.1"
                      /db_xref="GI:28828497"
                      /translation="MAANFWESSHCTEWLLDRNKIEESNPKDKQYLTPIELKKLRTHY
                      CFVIQNLGNALKLRQRAISTAIVYFKRFYLKNSFVDCEPRLVAVTCLYLSSKVEECIT
                      QAKKCAAKMKEIDHSFNYLMNDILECEFFVLEELDFCLIIYHPYKSLPFYLQSSGLDP
                      ASIEIIWGIVNDSYRTDVCLLYPPFVVGLGCILLGSYLLKKDIKQWLSELNVEMKDIW
                      EVSKDLIDYYEFEKQQSLQNQSELDLLYNKLPIRNKK"
     CDS              complement(join(3639. .4158,4352. .4422))
                      /note="GeneID exon scores (in order of location ranges):
                      63.23, 9.00 - GSCJ_ID dd_02845"
                      /codon_start=1
                      /product="similar to Dictyostelium discoideum (Slime
                      mold). Vacuolar ATP synthase proteolipid subunit (EC
                      3.6.1.34)"
                      /protein_id="AAO51106.1"
                      /db_xref="GI:28828498"
                      /translation="MFQLLSFLLSGEATAVERIITDACPVYAPFFGAMGVTAALVFTV
                      MGAAYGTAKASVGISNMGVMKPDLVIKAFIPVIFAGVIAIYGLIICVILVGGIKPNAN
                      YTLMKSFTDLGAGLTVGLCGLAAGMAIGIVGDSGVRAFGQQPKLYVIMMLILIFSEAL
                      GLYGLIIGILLSSVSDTYCPGQALVPLNSGNVIGKN"
     CDS              join(5902. .6004,6157. .6199,6351. .6486,6853. .6962,
                      7066. .7249)
                      /note="GeneID exon scores (in order of location ranges):
                      6.21, -0.71, 4.93, 17.98, 20.83 - GSCJ_ID dd_02846"
                      /codon_start=1
                      /product="similar to Mus musculus (Mouse). Similar to
                      ATPase, H+ transporting, lysosomal (Vacuolar proton pump)
                      21kD"
                      /protein_id="AAO51107.1"
                      /db_xref="GI:28828499"
                      /translation="MPDINPLTPIPDNNYYDQCTYFLVTISPSTWAALGIGLSLALSV
                      VGSAWGIWVTASSLMGAAVKEPRIRSKNIISIIFCEAVAIYGIILAIILNGKIDKFLN
                      IWDPASDYMAGYMMFGAGITVGLCNVFSGVCVGIAGSGCALGDAQNPSLFVKMLIIEI
                      FAGALGLYAVIVGILMTTNALGVTQQPALKP"
     CDS              complement(7573. .8283)
                      /note="GeneID exon scores (in order of location ranges):
                      66.20 - GSCJ_ID dd_02847"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO51108.1"
                      /db_xref="GI:28828500"
                      /translation="MSTTIYFSYPTYSNEREFTTYDDDSEEEYYDQSYFQDYSFQNKP
                      QPSSSNINKSKSVKVTGYGNGSKNAGQAIPRTDKDYKAAIEYKKEKKLAIKQIKENKL
                      STSPTTITTTTTTNTATSYENVNIESPNDENNKTKKLTKLEVKSIIEKETDLKKKERL
                      PSSIEVYNRILHDEQLKDKIQDFIITYEDRFDGLCDSELLKLDEAIPFHRIRQIRYLN
                      QIVWSRSERFYDFSNLLD"
     CDS              complement(join(9480. .10271,10379. .11125,11253. .11858,
                      11963. .12121))
                      /note="GeneID exon scores (in order of location ranges):
                      50.34, 56.92, 38.65, -2.17 - GSCJ_ID dd_03351"
                      /codon_start=1
                      /product="similar to Plasmodium falciparum (isolate 3D7).
                      Hypothetical protein"
                      /protein_id="AAO51109.1"
                      /db_xref="GI:28828501"
                      /translation="MEDTALMGYFHFIRYLHAATTPKLRFAFEIFSLIVISLFVSTLF
                      VMHNQFVGKSDCLKLPNNTRFDILSIEIKGETSIFSRQHIYKKSTTTTETTPTTTTTT
                      TTTVNNDIDQQSYNKYLQVLKGAVDKHEPCIYDENYDNDNDNDIDKDSVTTATTTTAA
                      AAAAAAAATTTTTTTSLITESTYFKTAIEFIIDLIRQMYHSLYNFISIFFNSQFDEYD
                      QQYKKDRLEREHLKRLKEELKKHIKDIDRNHVNFEKLQHHQQQQLQQPQQQQQQQQQQ
                      QQQQQQQQQQQQQQQQEREEKKEKDKQHNNNHHNNGIGSGVLNDLNDIVQYFIDSHNP
                      RYEFSYQNAFLMLTTEQRNKFNITTLHIDINANDQCLGNGFHKAIIGMLGYDILIINN
                      LVYGFPLIPNNNINNNNNNNNGRGGYLRIAGNSQLINLYQYSSPKLSLNNESILYIGY
                      RFFETNFIFFLTLIVFGSSLKLCESSFAHYVDRDPRGVLKCMLHNLLIFALFIFIQIG
                      YISIFSIILSTEYLVFFISSLITSYFSTWGLRSKESIRYYPIIFLSILYGLILYVVFF
                      PSGFHPIAFYACYSLIEYLLIICLFNFEIPAVLNNRILPNNQNNNNQNNNNQNNNNNN
                      QNNNQNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNQ
                      NNNQNNNNNQNNNNNNTNINNHNNNNINNQNNINNTQNENQNNNNQNDGNNQNIYNVP
                      INQQTIVNNSSQRRLAASAILNNSNEM"
     CDS              complement(join(13033. .14475,14560. .15048))
                      /note="GeneID exon scores (in order of location ranges):
                      138.82, 42.31 - GSCJ_ID dd_00640"
                      /codon_start=1
```

```
                     /product="hypothetical protein"
                     /protein_id="AAO51110.1"
                     /db_xref="GI:28828502"
                     /translation="MNIINYFKNGNNNNGHSSGVGSSGHSYYSNNKGTSFVPSPVLYM
                     TPMAPIDYGCPDDMLDDGRSDTAPSSQNTSPMYQHHNFHNNHNNTHNHNNNNNNNNNN
                     HEQPNCALEFLEHHLSFETKYGVNVSQGNRKYQEDRHKVKMGLENNQYLSLFGVFDGH
                     GGDRASNFVKKKIVNCVNKYVKENKAGYSSKNLNSSSSPTGSTSSASSSGSSSSSASS
                     SGGGGGGGGPLNGSSSSIGMPHASTISPTPSSGNSSTSSGASSSGVSSANSPIGVGFD
                     SPGWMKFNHLYNQNNFQQEIQNRSEFLQAALYNTFQFLDNRYCKKYRQKGDGGTTCLV
                     ALLSNPPNAQPLLVVANAGDSRGVLCRNGKAYALSYDHKPGNPKEKQRITSSGGKIEW
                     DFNERIWRVSGILSVSRGIGDIPLKKWVICDPEFVVFPLKGPIRKSNSSRYTSPSSSS
                     SSSSSSSSSSSSSSPSSPKAFRSSPQINRFFFSPSFKNKTISNSTGSLPPPSTNNSPQ
                     QPPPILHQNNISSPVSSISEAQPFNSIDNQNNNNNNNNNSKLISTSPPNHKMYYCSNH
                     QYNYGEVDQFFVLATDGIWDVFENQELVEFINAIIEESYHSKRLDWDPNEISKRVVQE
                     AYRKGSGDNATVLIIKLYW"
CDS                  complement(join(17600. .18734,18858. .19218,19315. .20422))
                     /note="GeneID exon scores (in order of location ranges):
                     69.94, 30.45, 124.05 - GSCJ_ID dd_00639"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). cGMP-stimulated cGMP phosphodiesterase
                     (cGMP-stimulated cGMP-specific phosphodiesterase)"
                     /protein_id="AAO51111.1"
                     /db_xref="GI:28828503"
                     /translation="MMFSKRKSIRFNPEGDYTELPRGGYVVPSKLGGIQFGVPPETIK
                     DSMALKIDVPTIYVFPEELWDRKTGINAAEAEFPAYFNYFILKRKVSFVCTKEQEQRI
                     RIVFQETLLGPPEFNTGIIINSSSSTTDTSKTSPIKKQTSSSSPPLSPQQQQPPPPLV
                     KQPSQQQLEELATMDTCHYHHHHHHQEVNDNDNNNNTTTNNNNIEILEQQQQQQQQQQ
                     QQQDEDSTDVDEEFQKEFSSTFPRSEIPNLEKECKYLRTFNSVDELFDFILFDDNGIA
                     KLSDDVEIHFQEDQSLFKVLQFEETGKGNKVQTLVATIPSKILFPDMINSISTINDNK
                     IFDPPTFGITIIGSSHGFDPKKSTTGFVLWINKRGIMVDPPLNSSSFLQSQGVPTRMV
                     DHIILTHCHADHDSGTFQKLLEEYQITVVTTPTILGSFLRKYGALSNLSTDLLRRLFI
                     FRPVMIGEPMIISGAEFRFFYTIHTIPTISFEVFYGGKSIFYSGDTCYDPNRIKDMNK
                     RGIMKTSRMKFFLRPWNHTVVLHEAGVPPIHTPVSVLRALPDEVKNRLYLVHISEHTL
                     PAGSGLKIAKEGVAHTLSLDVMKSSHSEAVDILKLVESVDIFRSIPLTQACEILQTAT
                     KRKYSQGSVIIARDTEPDAFYVVASGVVCVNIGELKKNLIVGDYFGEMSLVMGGLRSA
                     NVQAVTDVEVLSFNKEDFLSITRNSTESIQFITRLWEMRNEKSWETMSLNSVFSRCTN
                     SQKTAIQSILVRELIKKDETLWCKGEEALFGCLVAEGSFVFKEDDSLESFSQGSFLGD
                     INAMTTQPPSIHKTTVVAKEESVIYKVLSQDLIKFFSNNPGIQLAFLDTIFVDALRDQ
                     VQQLVNSKLTY"
CDS                  complement(21566. .21865)
                     /note="GeneID exon scores (in order of location ranges):
                     24.15 - GSCJ_ID dd_00638"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAO51112.1"
                     /db_xref="GI:28828504"
                     /translation="MTQLGYQQIPPGMHNNNNNNPNNNSNSSNNNNNSNFIDFLRQSG
                     ITMNDDVNSSGMMTPSSLLNDSSLEHRVASTNKELKDLEDKLQKSSPLNTQSYYR"
CDS                  complement(join(21982. .25164,25267. .25396,25496. .25821,
                     25923. .26082,26307. .26464,26602. .26691))
                     /note="GeneID exon scores (in order of location ranges):
                     340.18, 15.33, 31.36, 13.30, 15.31, 10.95 - GSCJ_ID
                     dd_00637"

  Query Match              5.3%;  Score 64.2;  DB 3;  Length 266544;
  Best Local Similarity   56.2%;  Pred. No. 0.051;
  Matches  140;  Conservative    0;  Mismatches  108;  Indels    1;  Gaps    1;

Qy        971 CTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGA 1030
              |||||  |   |||||  || |||  |   || |  || |||  |    |   ||   || |
Db     248420 CTTTTAACTTCAAATTGTTTACTAGAATGTAACGTAATCTCATCAGAACTTTTGGAAATA 248361

Qy       1031 ATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTG 1090
              |||||  |   |  || ||||  ||| ||  |  |   ||  || | | || | |   |
Db     248360 ATTAATGGTTTTAAATCTTATCATTTATCTATAAATAATCATTTAGATATTTTAAACTAT 248301

Qy       1091 AAAATAAATAAATTTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAA 1150
                || ||| |||    ||| |||   |      ||  | ||||   |  |  ||||  |||
Db     248300 TTAAAAAAAAAAAAACTTTAAAAAAAAATAAATAAAAATAATGTAATTATATGTATTAAA 248241

Qy       1151 TGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAA 1210
              | |   ||| ||| |||| || ||  | ||| || |   |||| |     ||||||||||
```

```
Db     248240 TTTAA-ATGATTTATTTAATTCTTTATATTTTATTTATAAAAAAAAAAAAAACAAAAAAA 248182

Qy       1211 AAAAAAAAA 1219
              ||||| |||
Db     248181 AAAAACAAA 248173


RESULT 4
AX655393
LOCUS       AX655393               2000 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 5263 from Patent WO03000898.
ACCESSION   AX655393
VERSION     AX655393.1  GI:29158207
KEYWORDS    .
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 5263 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1. .2000
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

  Query Match              5.3%;  Score 64;  DB 6;  Length 2000;
  Best Local Similarity   10.9%;  Pred. No. 0.15;
  Matches   89;  Conservative  362;  Mismatches  357;  Indels    8;  Gaps    2;

Qy        103 TTATTTACAAGGATACTGAATATGACTGCACTGAATATGCTAAATCAAATAAGCATCCTG 162
              :: |:: :   :| :| :|   |: :: | :: :   ::  :::::::  :::  :  ::
Db        220 KYGTYKKMTYYSASRCMRAYMTTSYSWACSSYTWCRSKRRSMMWKMMRKMRWSRSYGWYS 279

Qy        163 GCGGTCTTAATTTCCTCAATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTTCA 222
               : :     |::::::::: :       : |||:   :|:: ::         :   :: :
Db        280 WSYKMWMCTAYKKSYYSRWCYMYRGGGWRGATRYWGRGYMSRMAMMYKKMYWYRGYKGMK 339

Qy        223 GAACACTCCATTCTAAGTAGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACTGGCGCAA 282
              : : |  :::  : ::::|  :  ::::::|   ::| ::: |:  : :::| :: :: :
Db        340 RGWWAGRMMMRSMCRWSKACYYMRWRMWRMTRRRRWAKKSSRTSRRKKRKWCMRKRKYKR 399

Qy        283 AA---TAAGAGGAGACTGAATCTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTTAAGC 339
              ::    ::: ::|::::: :       :  :|  ||  :  :   :: :: ::|: :::::
Db        400 MRGYSRMRSCKRARWMKRCRSGRAWKMGCRGCMTCRMKSYGMMRWKSWKRMASKYKWMSR 459

Qy        340 ATTTATTCGAACCAAACTGGCCTATCGAAATTGGTTTATTCTTAACTACCTTTACTTTAT 399
              :: ::::   :  ::  |:| :  |:|: :::::     :|  ::  :::   :  :::
Db        460 MYRWRKKKCSRTTMWGKTRGGMMGTMGRCRYKKRSGMKRKCRRRRWGRMYRMRWKRYYMS 519

Qy        400 TTGTCACTGGATGTTTGACTCAAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTAATGC 459
                |:::  :  :  :  | : : : : :|: :  ::::::: :::::: : :  :  : :
Db        520 ARYTMRYCARKKYSYSAARKARCWYRGKGYYWAGMWMKRYKRMYMYKMMWWYKRKYSKCS 579

Qy        460 AAATCATCAGTGGTTGGATTGGTCACTCTATGAACCACAATCGTAACCCTATATTAAGAA 519
              :  :: ::|:  :  ::|  :   :::: |::  :  ::   :  |   : |::: ::| :
Db        580 WYCKMSYYASCMKSARKAGAKMCKRSKMSAWSKSMRSSRKCRKCASKRSSAKRYAMMGGM 639

Qy        520 AATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAGGA 579
                 : ::  ||:: :  ::::| |     :::|: ::   :: ::: :    ::
Db        640 TSGSRMSRWKSYTCYWRKWGSMKSTCTWMYYMSKYTYAKYGSYWRYRYRAWCMYMWRWYY 699

Qy        580 AGCACAATCAACATC-----ATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTA 634
               :::: :|::|  |:     :|  ::  :::     : : ::  :: :: :  ::|: :
Db        700 YRYRSYMTYMAWYTSSTRMAMTGMKYSGRYWTSWYKYCKCSWKYRSMWYYWSWWWAKTWK 759

Qy        635 ACACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCAT 694
```

```
                : :  ||  ::::::    ::  :  :  :: ::: |: :::  : :   |:: ::  ::|
Db        760 MWRRYATRMMWMWYRYSMKWYTWCTMWGYWWYWWRTYMKMRYMWYKCTKTYWYWSATYWT 819

Qy        695 CTTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATT 754
               |:|  :  :|  :     :|: : : | :::   | | |:: : :|:|  : :|:: :
Db        820 GTWAAWWMAKTKMRMGMTGAKTRGRARKARYWWKWATWCATKRWMTKGKGAKWAWTWMAK 879

Qy        755 CAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTC 814
               :: : : ::   : : : :  ::: ::  ::  ::  :  :   ::|  ::::::: | :
Db        880 AWRKYYWSWMRAWYYYYKTRRTRYKTCWWKARWGSWAYWRMWWKGSAKMWWMWKGGRWGW 939

Qy        815 TATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATT 874
               | ::::: || ::|   ||: :   |::|     :|   ::: ::::: : |   :|
Db        940 TKYWYWYCTTWKMACGRATKYMCCAGWWAMYSYSWTRTYWMRTWRWMWASSRTAKRMARM 999

Qy        875 TTTCGAACATCAAATAGCTGCAAGCAGAAACTACGC 910
               :: :|::  :|:|:::: |||   | | ||||| |
Db       1000 MWKTRAWSKSYARAYWKMAGCACCTACACACTACAC 1035


RESULT 5
AE014818/c
LOCUS       AE014818               252632 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 14 section 3 of 13 of the
            complete sequence.
ACCESSION   AE014818 AE014187
VERSION     AE014818.1  GI:23497188
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 252632)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 252632)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .252632
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="14"
     repeat_region   76. .110
                     /rpt_type=tandem
     repeat_region   140. .166
                     /rpt_type=tandem
     repeat_region   828. .857
                     /rpt_type=tandem
     repeat_region   1548. .1607
                     /rpt_type=tandem
     repeat_region   1846. .1913
                     /rpt_type=tandem
     repeat_region   2010. .2087
                     /rpt_type=tandem
     repeat_region   2370. .2401
                     /rpt_type=tandem
     repeat_region   2560. .2613
                     /rpt_type=tandem
     repeat_region   2567. .2627
                     /rpt_type=tandem
     repeat_region   2930. .2964
                     /rpt_type=tandem
     gene            <3001. .>4509
                     /locus_tag="PF14_0124"
     mRNA            join(<3001. .3448,3827. .>4509)
                     /locus_tag="PF14_0124"
     CDS             join(3001. .3448,3827. .4509)
                     /locus_tag="PF14_0124"
                     /codon_start=1
                     /product="actin II"
                     /protein_id="AAN36736.1"
                     /db_xref="GI:23497189"
                     /translation="MSEEAVALVVDNGSGMVKSGLAGDDAPKCVFPSIVGRPKMPNIM
                     IGMEQKECYVGDEAQNKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVSPEEHP
                     VLLTEAPLNPKTNREKMTQIMFETFDVPAMYVSIQAILSLYASGRTTGIVLDSGDGVS
                     HTVPIYEGYVLPHAINRIDMAGRDLTYHMMKLFTERGHTFTTTAEREIVRDIKEKLCY
                     IAMDYDEELKRSEEHSDEIEEIYELPDGNLITVGSERFRCPEALFNPTLIGRECPGLH
                     ITAYQSIMKCDIDIRKELYNNIVLSGGTTMYNNIGERLTKEMTNLAPSSMKIKVIAPP
                     ERKYSVWIGGSILSSLSTFQQMWITKEEYEDSGPSIVHRKCF"
     repeat_region   3580. .3605
                     /rpt_type=tandem
     repeat_region   4965. .5005
                     /rpt_type=tandem
     repeat_region   5181. .5210
                     /rpt_type=tandem
     gene            complement(<5327. .>6817)
                     /locus_tag="PF14_0125"
     mRNA            complement(<5327. .>6817)
                     /locus_tag="PF14_0125"
     CDS             complement(5327. .6817)
                     /locus_tag="PF14_0125"
                     /codon_start=1
                     /product="deoxyhypusine synthase"
                     /protein_id="AAN36737.1"
                     /db_xref="GI:23497190"
                     /translation="MVDHVSFKEVNKIRSDDECDADSHNEGDNIEDAKASVFVKSSLI
                     PEKTDVVKGLNFDKEVDLHEFINNYKYMGFQATNLGISIDEINKMIYYKYKDENIKSE
                     PNNENNLNCNNVSEDLNKDQENHLYHYEKKKKSCIIWLSFTSNMISSGLREIFVYLAK
                     NKFIDVVVTTAGGIEEDIIKCFSNTYIGDFNLNGKKLRKKGWNRIGNLIVPNDNYCKF
                     EDWLQPILNKMLHEQNEKNEQMFLKKLEKRKKKYNNNKNKNDNNNDNDNVWGNEKNDQ
                     NQNQYNQGQESFKKDSNIYTNDVSNKKNHINNYINNYDSDSDDQCDMYYLSPSEFINT
                     LGKEINDESSLIYWCYKNDIPIFCPGLTDGSLGDNLFFHNYGKKIKNNLILDIVKDIK
                     KINSLAMNCEKSGIIILGGGLPKHHVCNANLMRNGADFAVYVNTASEYDGSDSGANTT
                     EALSWGKIKYGQTNNHVKVFGDATILFPLMVLNSFYLYDQKRKKDM"
     repeat_region   5879. .5904
                     /rpt_type=tandem
     repeat_region   6852. .6887
                     /rpt_type=tandem
     repeat_region   7054. .7079
                     /rpt_type=tandem
     repeat_region   7268. .7298
                     /rpt_type=tandem
     repeat_region   7619. .7663
                     /rpt_type=tandem
     repeat_region   7838. .7887
                     /rpt_type=tandem
     gene            <8004. .>11663
                     /locus_tag="PF14_0126"
     mRNA            <8004. .>11663
                     /locus_tag="PF14_0126"
     CDS             8004. .11663
                     /locus_tag="PF14_0126"
                     /codon_start=1
                     /product="hypothetical protein, conserved"
                     /protein_id="AAN36738.1"
                     /db_xref="GI:23497191"
                     /translation="MNINFYVYDDEADNEIISSNFNVYCSDEFLKIHNIEKNDNAIIY
                     HEIMRQDEEEGQEKGQEIGQEKDQEKDQKKDQNKDQKKDQNKDQKKEQARNINHINNI
                     HININKIDDEYVEKEPKPNVQNFSVLCCFKNYNELNIKKKLENNGLCINKYILKSLNL
                     DIKEIEEKNIYKSCKDQNNNMKKIGIFKVSIIKPQLFIPIKEIKLHVQEIFNTSYEYL
                     KSYQYENIDLLQSFWKNIKSNNMEKYINLSLLNTCLFVNNFLKITILGVHSYITVGQI
```

```
                     ILTKDKKTKMKNNNNNNKNKTVRNMNARDKQENNKNGDNKNKNGDNNNKNGDNKNKNG
                     DNKNKNGDNKNKNGDNKNKNGDNNNNIVVPFIDQNTKIVIISNDDIKNVVDDKHIINK
                     YDKEKTNINSDHMKKLNLKKMVVNKKEKKGLNKLGGYNKIKEDIYYYILIPLIYKHIY
                     DAYHIDIHKGILLHGPPGCGKTYIALLIKEELSLLKKKFCTITQNKTSKQLNNTKNDN
                     NYNNDDNIIIPEIEILKSTDLIDYNNTGYKINELFLRCHKRYQEENKCTIIFIDEIEI
                     LCKKREENNNMNIYTSVLLNNMDGIKKHTHTILIGATNYINKIDLALRRSGRFDKEIE
                     VNLPNLKDRISIFQKKLNLIKHNIGKKKIHKLADLCQSFTCSDINSLINISMFLNLKE
                     NKIISRDIFKTANVHTIDDTNDIKNDQTSACLEDCIEGKIGTNEMHNTKENKYINMNN
                     NNDNDNNNDNNNDNNNDNNNGDNNNYYLTPYDKNHLYNNSLINTQSNSCHEKKMMNYN
                     KNNLYLLRYKHLLKALKYVKPSGMKELYIDIPKTRFKDIGGYKFVKRCINECLIYPKK
                     YKNIYDKYNIESPKGILLYGPPGCSKTLFAKAIASEIHMNFISVKGPEIFSKYVGESE
                     KSIRNIFKKARENHPCVIFFDEIDSIAVNRNNNQNFVSNRVLCQLLNEIDGIKNRLNV
                     IILAATNRPDLIDPALMRPGRFDRIIYVPLPNYSSRFAILKKNLKFFKIHNLIEYDKK
                     ETINDLNHNEIKRECIIQDIQYNDHPKNSNQKINNHITKSSYDHIHNYTFNKNNMNIK
                     TNEDSSLYKHYKYTDIKFIKLCHVLAKKTKKYSGAEIVNICREASICALRQTLKKRNI
                     KINPKIKIKIKNNETNNTQNNIAHSEKNKYPFDKEPLIGLSKKHFLRVLKKIKPQTSE
                     SLINFYKNYKEQNKV"
repeat_region        8158. .8214
                     /rpt_type=tandem
repeat_region        8197. .8275
                     /rpt_type=tandem
repeat_region        8931. .9082
                     /rpt_type=tandem
repeat_region        10221. .10298
                     /rpt_type=tandem
repeat_region        10225. .10295
                     /rpt_type=tandem
repeat_region        11929. .11981
                     /rpt_type=tandem
repeat_region        12161. .12194
                     /rpt_type=tandem
gene                 complement(<12215. .>13834)
                     /locus_tag="PF14_0127"
mRNA                 complement(join(<12215. .13261,13406. .13453,13580. .13702,
                     13820. .>13834))
                     /locus_tag="PF14_0127"
CDS                  complement(join(12215. .13261,13406. .13453,13580. .13702,
                     13820. .13834))
                     /locus_tag="PF14_0127"
                     /codon_start=1
                     /product="N-myristoyltransferase"
                     /protein_id="AAN36739.1"
                     /db_xref="GI:23497192"
                     /translation="MNDDKKDFVGRDLYQLIRNAKDKIKIDYKFWYTQPVPKINDEFD
                     ENVNEPFISDNKVEDVRKEEYKLPSGYAWCVCDITKENDRSDIYNLLTDNYVEDDDNV
                     FRFNYSSEFLLWALSSPNYVKNWHIGVKYESTNKLVGFISAIPIDMCVNKNIIKMAEV
                     NFLCVHKSLRSKRLAPVLIKEITRRINLESIWQAIYTAGVYLPKPISTARYFHRSINV
                     KKLIEIGFSCLNTRLTMSRAIKLYRIDDTLNIKNLRLMKKKDIDGLQKLLNEHLKQYN
                     LHAIFSKEDVAHWFTPIDQVIYTYVNEENGEIKDLISFYSLPSKVLGNNKYNILNAAF
                     SFYNITTTTTFKNLIQDAICLAKRNNFDVFNALEVMDNYSVFQDLKFGEGDGSLKYYL
                     YNWKCASCHPSKIGIVLL"
repeat_region        13286. .13345
                     /rpt_type=tandem
repeat_region        13525. .13560
                     /rpt_type=tandem
repeat_region        13736. .13793
                     /rpt_type=tandem
repeat_region        13834. .13865
                     /rpt_type=tandem
repeat_region        14004. .14035
                     /rpt_type=tandem
repeat_region        14045. .14069
                     /rpt_type=tandem
repeat_region        14472. .14498
                     /rpt_type=tandem
repeat_region        14506. .14540
                     /rpt_type=tandem
repeat_region        15062. .15088
                     /rpt_type=tandem
repeat_region        15090. .15121
                     /rpt_type=tandem
repeat_region        15279. .15356
                     /rpt_type=tandem
repeat_region        15279. .15354
                     /rpt_type=tandem
repeat_region        15666. .15709
```

```
  Query Match              5.1%;  Score 62.4;  DB 3;  Length 252632;
  Best Local Similarity   47.7%;  Pred. No. 0.11;
  Matches  253;  Conservative    0;  Mismatches  266;  Indels   11;  Gaps    2;

Qy        696 TTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTC 755
              ||  ||| || || |   ||| ||  | ||||   |    || | ||   |||  ||||
Db      70032 TTTCCTTTTTTTTTTTTTTTTAAATATTCTTTTGTATATTTCTCTAAATATATTTATAT 69973

Qy        756 AACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTT------CAGT 809
              |  || |  ||         ||  |  || |  | | ||   ||  | |||      |  |
Db      69972 ATATATATATTAATATATTTATATAATTATATAATTATATATGTATACTTCCTTTCCTTT 69913

Qy        810 GCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTA 869
              |||  |||| |  ||   | |  | ||||| | || |||||  |||||    | | |||
Db      69912 CCTTTTATTTTATTTTATATTTCTTAAAATTACATAAAATTTAAAAGATTGTTAATTTTT 69853

Qy        870 CCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCT 929
              ||  ||  |||| ||  ||||    |||  | |   |   |||    ||||| |
Db      69852 GAATATTACGAATATGTAATAATGTCAATTATATTTTCAATTTTAATTGACATGTA---- 69797

Qy        930 CTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCT 989
              ||  |   ||  || | |  || | || | | |   |    | | |  ||  ||
Db      69796 -TATGTGAAATATATGTTTTAAAAAAATATATTTATTTTTGTATGTGCATATTTATAATA 69738

Qy        990 TTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTT 1049
              || || | || |  || |      | | |||| |    |     ||||   |    |
Db      69737 TTATATATATAATACATATAATATTTTATAAATATGTATTTTATAAAATGGAAACATAAA 69678

Qy       1050 AAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTT 1109
              || |   | ||  | |  |||  || |  |||   |   ||| |  | |  |||||
Db      69677 AAAAAAAAAAAAAAAAAAAAATTTTTAACAAATCAATATATTAAAAAGGTTATATTATTA 69618

Qy       1110 TAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTAC 1169
              |||||  |||||          |||   |||  | ||      | | || |  ||| | | |
Db      69617 TAAATAAATATTACGAATAATATATATACATATATATATTTTTTTTTTTTTTTTCTTC 69558

Qy       1170 TTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219
              ||  |  | | |  |||    ||   |   ||    ||  |  ||  |||
Db      69557 TTTCTTTTCTGTTTTCTTTATAATGGGAAATATTTTAATCATCAATTAAA 69508
```

```
RESULT 6
AX599046
LOCUS       AX599046               8056 bp    DNA     linear   PAT 14-FEB-2003
DEFINITION  Sequence 386 from Patent WO02077272.
ACCESSION   AX599046
VERSION     AX599046.1  GI:28399186
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Berlin,K., Braun,A., Distler,J., Guetig,D., Howe,A., Mueller,J.,
            Olek,A., Piepenbrock,C., Adorjan,P., Grabs,G., Lesche,R., Leu,E.,
            Lewin,A., Lipscher,E., Maier,S., Model,F., Mueller,V., Otto,T.,
            Pelet,C. and Ziebarth,H.
  TITLE     Methods and nucleic acids for the analysis of hematopoietic cell
            proliferative disorders
  JOURNAL   Patent: WO 02077272-A 386 03-OCT-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .8056
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match              4.8%;  Score 59;  DB 6;  Length 8056;
```

```
  Best Local Similarity   44.3%;  Pred. No. 0.83;
  Matches  483;  Conservative    0;  Mismatches  585;  Indels   23;  Gaps    5;

Qy        101 ATTTATTTACAAGGATACTGAATATGACTGCACTGAATATGCTAAATCAAATAAGCATCC 160
              |||| || | ||  | |   || || | |     |  | |  || | |  |     |
Db       2396 ATTTTTTGAAAATTAAAAAAAAAAATAATTTATTTATTTTTAAAAATTAATTTTTTTTTAT 2455

Qy        161 TGGCGGTCTTAATTTCCTCAATTTGTTTATTGATGAGAAGTAAGATTTGACTGAATATTT 220
              |      |  ||  ||  | | | | ||  || |  | || ||| || | | |||| | |
Db       2456 TTTTATTTATATATTTTTTATTATTTTGTTTTAAAATAATTAAAATATTAATGAAAAATA 2515

Qy        221 CAGA-----ACACTCCATTCTAAGTAGGCTTTGAAAATTTTAAAATCCTTCCCTAAGACT 275
               | |     | | |  ||| |||       ||  ||||||| |||||         ||   |
Db       2516 AAAAAAAATAAATTAAATTATAAAATAAATTAAAAAATTTAAAAATTTAAATAAAATTTT 2575

Qy        276 GGCGCAAAATAAGAGGAGACTGAATCTTCAAAGAGATTCTCAATATTAAAGAAAAAGCTT 335
                || | ||  |  |     ||      |     | | | |||| |  | |||||
Db       2576 TAAAAAATAAAAAATAAATGAAAAAAAAAAATTATAAATATAAATAATTTAAAAAAAAAAA 2635

Qy        336 AAGCATTTATTCGAACCAAACTGGCCTATCGAAATTGGTTTATTCTTAACTACCTTTACT 395
              |   |   |   |    ||  |    | |  ||| |  |  ||| |||| |   ||||
Db       2636 ATAAAAAAAAATGTTAAAATTTAAAATTTTAAAAATAATAAATTTTTAAATTTTTTTAAA 2695

Qy        396 TTATTTGTCACTGGATGTTTGACTCAAAAGTGGTATTTCTCTATTCCCCTTCTTGTCTTA 455
              ||  || |   |   | ||       || | |     |   | ||           | |||
Db       2696 TTTATTATTTTTTTTTTTTAATAAAAATAATTTGTTAAATAAAT----------TATTA 2744

Qy        456 ATGCAAATCATCAGTGGTTGGATTGGTCACTCTATGAACCACAATCGTAACCCTATATTA 515
              ||  |||| |  |     |  | | ||   | | | |   |      | |   |   |||
Db       2745 ATAAAAATAAAAAAAATATTTAATTGTTTATTTTTAATAAATTTAAATGAATTTTGTTTA 2804

Qy        516 AGAAAATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGT 575
              ||||||    ||| |      |  | ||| |           ||| || | |
Db       2805 TTAAAATTGTTATTAATAATAAATGATATTTTTTTATAAAATATTAAAAATTTGAAATTA 2864

Qy        576 AGGAAGCACAATCAACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAA 635
                | | ||  |  | |      |  || ||  ||  ||| ||  || | || | | |
Db       2865 TTTTATGAAAAAAATTTTGTTGAAAAAAATAAAAAATTTGAAAATGAAAAAAAAATGTTA 2924

Qy        636 CACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATC 695
               |   |||  || |||   | ||       |||||  ||       ||||  | |    ||
Db       2925 AAATTTTATGAAATGTAAAATTTAAATGAATTATTATTGTTAAAAAAATAAAAAAAAATT 2984

Qy        696 TTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTT---GCACTGGGTATTATTA 752
               | |      |      | ||| ||   ||  |  ||   ||     | |    ||||  |
Db       2985 ATTGAAAAATTGATATATTTTAAAATGATGATGTTTAAATTTTATAAATATAAATTAAAA 3044

Qy        753 TTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCT 812
               | || ||||  || ||  |   |||    | |||||| |||| |  |||  |  ||  |
Db       3045 ATAAAATAAATGTTATAATTGTGAATGAAATTTGAATT-ATTGTTATTTTTATGTGTAAT 3103

Qy        813 TCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCA 872
              | ||      |  |  |  | |||||||| | ||    ||  ||| |||   ||||
Db       3104 TTAATAAAAAATTAATTTTTTTAAAATGAATTTAATGATGATTGAAAAATTGTTTTATAT 3163

Qy        873 TTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTA 932
               |  |   | || | |    |   | | ||| |  |. | |     ||  |   |  |
Db       3164 ATAATTTTAAATAATAATATTTTTAATAAAAATTTGTTGT---TTGAATTGTAAATAAA 3220

Qy        933 CTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTC 992
               || | ||   |  ||| |  | |||        || | | |||   || ||||  ||
Db       3221 ATTTTAATTTATTATATTTTTTTTTAATAATTTGATTATTATTTTTAATTAATTGATTGT 3280

Qy        993 TACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAG 1052
               | | | |        |  |||  | |||  |   || ||| || |  ||   |||
Db       3281 AAAAAAATTTTTGTTATTGTTGTTGTTATTTTTTTGAAAAAAAAAAAAATATAATTGTAAT 3340

Qy       1053 ATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAA 1112
               |  | || ||   |  |||||||          | || ||  ||||  ||  |||
Db       3341 GTAAAAAAATGAATAAAATTTGAAATGATGATTAAAAAATAAAAATGAAATAAATAATAA 3400

Qy       1113 ATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTA 1172
                | |  |||| |||   ||  |   |      ||||  | | ||   || | |  |||
```

```
Db       3401 TTTGAATTTTTTTTAAATATTTTTGTTAAATGTTGAAAATTATAATTTGTTTTATTATTA 3460

Qy       1173 TTACTTTTTAA 1183
                 |||  |  | ||
Db       3461 TTATTAATAAA 3471


RESULT 7
AX083744
LOCUS       AX083744                1141 bp    DNA     linear   PAT 28-FEB-2001
DEFINITION  Sequence 22 from Patent WO0111061.
ACCESSION   AX083744
VERSION     AX083744.1  GI:13185472
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Kunst,L. and Clemens,S.
  TITLE     Regulation of embryonic transcription in plants
  JOURNAL   Patent: WO 0111061-A 22 15-FEB-2001;
            UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES             Location/Qualifiers
     source          1. .1141
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
     promoter        1. .1141
                     /note="consensus sequence of A.t., L.a., and B.n. FAE1
                     promoters"
ORIGIN

  Query Match             4.8%;  Score 58.4;  DB 6;  Length 1141;
  Best Local Similarity   11.9%;  Pred. No. 1.6;
  Matches   83;  Conservative  261;  Mismatches  352;  Indels    4;  Gaps    1;

Qy        518 AAAATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTTTCTCTAATAAATGGTGGGGTAG 577
              :::| :  :::: :|       | : : ::::::|: :::     |  :::: ::       |
Db        129 RWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRTRWWWKKNNNATGWDDDTKYHMWNNNG 188

Qy        578 GAAGCACAATCAACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACA 637
                : ::::  : ::: ::  : :  :      :|:::::::::|:: :: :  |: :
Db        189 CBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYDVTYYWWVWDDMCKRKVRRWVRTRGRM 248

Qy        638 CGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTT 697
                  |  |: :     :|: : :  |       | |  ::|  : : : ::  ||||
Db        249 RNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNNNNNAKAMCKRAKYWGWNRABVNSTCTT 308

Qy        698 AGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAA 757
              ::::| : :|  :       |:| :::: :: :    :::         | ::| :    ::
Db        309 WKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAAMGVYWNNNNNNNWTYKKARHBARWDW 368

Qy        758 CTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTAT 817
              ::: |: :::   |:::  :|::: | :     ||  :: : ::::::: : : |:
Db        369 VWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNN 428

Qy        818 TCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTT 877
                :|: |:           ::|   :::|:::         :: : |::  | ::  |
Db        429 NNGGRTYYGWTKNKKMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMT 488

Qy        878 CGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTAT 937
                | ::: ||::  :   ::::| |    :  ::  | ::: ::  ::   :  ::|:
Db        489 NGKSHRBAAAVYTWYMWWWRRYAHANNNNWDYWWKACTWYKYBVCSKWWNNYAAWYTKSS 548

Qy        938 TATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAG 997
              : : ::: :: |   : :  : :: |:|  : : ::: :     :: : : :::: | |
Db        549 WNYTSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTWBHBRAGAA 608

Qy        998 ATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCA 1057
              :::::::  |:|:| : ::    : : :  ||                | |:::::|
Db        609 HYWMBMMYBAKCHCMKAWY----KAKKYAGAGGSNNNNNNNNNNNNNNNNATCARDDYYA 664

Qy       1058 TGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCA 1117
```

```
                :::  :    ::::::||  |::|:|  ::  :  ||: :  :::         ::
Db        665 ASRWYAMANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWKNNNNNNAGTWKNNN 724

Qy       1118 TATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACT 1177
                : |  |:  | |   |   ::::| ::|: |:  :|  :||::     :| :|
Db        725 NNNAKNASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADAAABTTDKRNNGAYTKYTTTN 784

Qy       1178 TTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAA 1217
                | : :    ||:: :                : ::::
Db        785 NNNTYRGVVTNTAARDGWANNNNNNNNNNNNNNGWSDMWV 824


RESULT 8
AX346592/c
LOCUS       AX346592               6103 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 1663 from Patent WO0200928.
ACCESSION   AX346592
VERSION     AX346592.1  GI:18494478
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 1663 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .6103
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             4.8%;  Score 58;  DB 6;  Length 6103;
  Best Local Similarity   46.6%;  Pred. No. 1.3;
  Matches  221;  Conservative    0;  Mismatches  250;  Indels    3;  Gaps    1;

Qy        748 TATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCA 807
              | |  || ||  |||| |  || | | |     ||| |   | | ||   || | |
Db       1844 TCTACTTAAAACAAAATCTAACTACCTTTAAAACATCTCTACAAAATCCTTTACTCCTAA 1785

Qy        808 GTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAAT---GAAATTCGAAAGAAGAATCA 864
                || ||| | || |   |||  |    ||  || |    ||||   ||  || ||  |
Db       1784 AAACTACTAATTTTATCCAAAACAAAATTAACCAACTATAAAAATCAAAACCAAAAAAAA 1725

Qy        865 CTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATAT 924
                | ||||   | ||   || |  |  ||    |||| |   ||   ||   |
Db       1724 ACAACTCCTTTTCAAATCACACAACAAATTAAAAACTAAACAAAAATTATAACTTAAATC 1665

Qy        925 TCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAA 984
              | |  |||| |  |||     |   ||    |  ||  ||  |||  | |  |||||
Db       1664 TTTTACTACATCGTATAATAACTTATTACATATCATCTCCACCACACCCTCTTTAATAAA 1605

Qy        985 TTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGA 1044
              | || |   ||| | ||| | || |||           |  | ||  ||| ||    |
Db       1604 TCCCATAAAACAAAATACTCCAATCTCCAACTCATTAAAATACATAAAAAAAAAACAATA 1545

Qy       1045 ACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATT 1104
              |    || |  ||       | | | |  | | |||    | | ||   |   ||  | |
Db       1544 AAAAAAATAAACACTTTTTAACAAATAATTATAAAAAAAAACCTTTCCATACAAACATAT 1485

Qy       1105 TATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTG 1164
               ||  ||||| ||| | |||| |  ||| |||   |   |  ||||    |||  |
Db       1484 AATAATAAATACATCTATTATAAAAAATCCTACATAAAAAAAAAAATTAAATATACTAAA 1425

Qy       1165 TTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1218
              |    | |||  | |         |  |||   | |||| ||| ||||| |||||
Db       1424 TACGTTCATTTTTCTAATCAAAAATTAAAACATATTAAAAAATAAAAACAAAAA 1371


RESULT 9
```

```
AE014834
LOCUS       AE014834               253001 bp    DNA     linear   INV 07-OCT-2002
DEFINITION  Plasmodium falciparum 3D7 chromosome 10 section 6 of 7 of the
            complete sequence.
ACCESSION   AE014834 AE014185
VERSION     AE014834.1  GI:23495173
KEYWORDS    .
SOURCE      Plasmodium falciparum 3D7
  ORGANISM  Plasmodium falciparum 3D7
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 253001)
  AUTHORS   Gardner,M.J., Hall,N., Fung,E., White,O., Berriman,M., Hyman,R.W.,
            Carlton,J.M., Pain,A., Nelson,K.E., Bowman,S., Paulsen,I.T.,
            James,K., Eisen,J.A., Rutherford,K., Salzberg,S.L., Craig,A.,
            Kyes,S., Chan,M.-S., Nene,V., Shallom,S.J., Suh,B., Peterson,J.,
            Angiuoli,S., Pertea,M., Allen,J., Selengut,J., Haft,D.,
            Mather,M.W., Vaidya,A.B., Martin,D.M.A., Fairlamb,A.H.,
            Fraunholz,M.J., Roos,D.S., Ralph,S.A., McFadden,G.I.,
            Cummings,L.M., Subramanian,G.M., Mungall,C., Venter,J.C.,
            Carucci,D.J., Hoffman,S.L., Newbold,C., Davis,R.W., Fraser,C.M. and
            Barrell,B.
  TITLE     Genome sequence of the human malaria parasite Plasmodium falciparum
  JOURNAL   Nature 419 (6906), 498-511 (2002)
   PUBMED   12368864
REFERENCE   2  (bases 1 to 253001)
  AUTHORS   Gardner,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .253001
                     /organism="Plasmodium falciparum 3D7"
                     /mol_type="genomic DNA"
                     /isolate="3D7"
                     /db_xref="taxon:36329"
                     /chromosome="10"
repeat_region        complement(79. .107)
                     /note="AT_rich"
                     /rpt_type=tandem
repeat_region        complement(113. .136)
                     /rpt_type=tandem
                     /rpt_unit="(A)n"
repeat_region        292. .441
                     /rpt_type=tandem
repeat_region        complement(292. .441)
                     /rpt_type=tandem
                     /rpt_unit="(TA)n"
repeat_region        455. .479
                     /rpt_type=tandem
repeat_region        complement(455. .479)
                     /rpt_type=tandem
                     /rpt_unit="(TAA)n"
gene                 <498. .>3257
                     /locus_tag="PF10_0307"
mRNA                 <498. .>3257
                     /locus_tag="PF10_0307"
CDS                  498. .3257
                     /locus_tag="PF10_0307"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAN35504.1"
                     /db_xref="GI:23495174"
                     /translation="MASCTGKVNFSSSGKEKGIGLESVLENAFDELISEFKKKFLHKI
                     IKVLEEAKYNENLDIMNERRKNFFLKKYENDQREDEDEYENKIGNVVEKKIEGNVEGN
                     VEGNVEREDIMIEEEKKKKKKKMDSSSSSNDSTFGVYLKSIIDRNKISLLKKLEKQKE
                     LKSVEENNKINYYNNKYKNGIIHYDENKEYHNNLNDKHVYNKNEESMNERTKLVKYTF
                     HDKNKNKNDMYLYPQESHFKYVSERIENLYKISNKLREEKETYHNLKYDKDNENQHFK
                     ENHTEDETDHEQNYEKYNVKLKNYVQNHNTQQMDKYQINEKNNISKLKQYNNNNDNNN
                     NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNLKGNMEKNIINNHQKNYISYYNNS
                     FKYPSRSHIFTERFSYHASDEHNEKSDHIYDMSPNKDKNCNNIYGILHNKDKHCNNNY
                     DMSSNKDKHCNNIYDMSPKKDKHCNNIYGILHNKDKHCNNIYGMSPKKDKHCNNIYGI
                     LHNKDKHCNNIYDMSPKKDKHCNNIYEMSPKKDKHCNNIYEMSPKKDKHCNNIYDNSH
                     LSSRGANKLPPSNNIIHEEKKQIINENINIESPQKINDLCNEKNNNLECKKEKIKLEK
```

```
                QFSFDSKDSDDNYVRGKNNMYEQDDLKNNMSQRYYKKVEDFKSYQLNKKISNEKEIIK
                RPNNNLFFEVIRGKKRKNMQGFECKDCKSFYNEIYSDDDEGKCVHKNFFMKNEQQLKK
                KININKDVRSKNSSTDDINSVDKMNNVDKMNNANNININNADNININNADNININNED
                NINNMDNLNNADNLNNADNLNNADNYPLNKKVHMNNSKKSIYYSHSLSPTCLNAHTDT
                EKASMNEKKNGNMLNMDNIKKENDLEKEKEKKKKKIIQTFSRHRYHNKINDSPKNFWK
                FDFFK"
repeat_region   781. .821
                /rpt_type=tandem
repeat_region   841. .870
                /rpt_type=tandem
                /rpt_unit="(A)n"
repeat_region   932. .1111
                /rpt_type=tandem
                /rpt_unit="(TAAAAA)n"
repeat_region   1156. .1175
                /rpt_type=tandem
                /rpt_unit="(TAAAAA)n"
repeat_region   complement(1369. .1397)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   1439. .1601
                /rpt_type=tandem
                /rpt_unit="(TAA)n"
repeat_region   1474. .1579
                /rpt_type=tandem
repeat_region   1479. .1588
                /rpt_type=tandem
repeat_region   complement(1618. .1669)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   1756. .2184
                /rpt_type=tandem
repeat_region   1775. .1954
                /rpt_type=tandem
                /rpt_unit="(TAA)n"
repeat_region   complement(2707. .2735)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   2772. .2814
                /rpt_type=tandem
repeat_region   2790. .2965
                /rpt_type=tandem
repeat_region   2808. .2966
                /rpt_type=tandem
repeat_region   3142. .3171
                /rpt_type=tandem
                /rpt_unit="(A)n"
repeat_region   complement(3274. .3347)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3391. .3437)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   complement(3439. .3503)
                /rpt_type=tandem
                /rpt_unit="(A)n"
repeat_region   3439. .3500
                /rpt_type=tandem
repeat_region   3440. .3500
                /rpt_type=tandem
repeat_region   complement(3561. .3683)
                /rpt_type=tandem
                /rpt_unit="(TATAA)n"
repeat_region   complement(3710. .3745)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3763. .3824)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   3763. .3811
                /rpt_type=tandem
repeat_region   complement(3871. .4025)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(3922. .3987)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   3926. .3987
                /rpt_type=tandem
repeat_region   complement(4037. .4082)
                /rpt_type=tandem
                /rpt_unit="(TATAA)n"
repeat_region   4043. .4082
                /rpt_type=tandem
repeat_region   4043. .4080
                /rpt_type=tandem
repeat_region   complement(4084. .4104)
                /rpt_type=tandem
                /rpt_unit="(A)n"
repeat_region   complement(4135. .4181)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   complement(4182. .4313)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   4364. .4466
                /rpt_type=tandem
repeat_region   4372. .4457
                /rpt_type=tandem
repeat_region   4377. .4472
                /rpt_type=tandem
                /rpt_unit="(TATAA)n"
repeat_region   complement(4512. .4655)
                /note="AT_rich"
                /rpt_type=tandem
repeat_region   complement(4670. .4775)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   4675. .4768
                /rpt_type=tandem
repeat_region   4870. .4912
                /rpt_type=tandem
repeat_region   complement(4870. .4912)
                /rpt_type=tandem
                /rpt_unit="(TA)n"
repeat_region   complement(4918. .5000)
                /rpt_type=tandem
                /rpt_unit="(TAAAAA)n"
gene            <5162. .>8973
                /locus_tag="PF10_0308"
mRNA            join(<5162. .5350,5488. .5535,5617. .5802,5902. .5981,
                6099. .6162,6254. .6349,6463. .6540,6727. .>8973)
                /locus_tag="PF10_0308"
CDS             join(5162. .5350,5488. .5535,5617. .5802,5902. .5981,
                6099. .6162,6254. .6349,6463. .6540,6727. .8973)
                /locus_tag="PF10_0308"
                /codon_start=1
                /product="hypothetical protein"

  Query Match             4.8%;  Score 58;  DB 3;  Length 253001;
  Best Local Similarity   46.8%;  Pred. No. 0.61;
  Matches  216;  Conservative    0;  Mismatches  245;  Indels    1;  Gaps    1;

Qy        758 CTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTAT 817
              || | |  | |||||  ||  | | | |  |    || ||  ||| ||   |  ||||||
Db     227329 CTGATATATATATATAATATATATTTACACAATTTCCTTTTTTTTTTTTTTTTTCTAT 227388

Qy        818 TCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTT 877
              | || ||  ||||   | |||   ||| ||  |   || ||     | |||    |   |
Db     227389 TTTTTTTACAAATACAGTAAAATTAATAAATATTCGAAAAAAGTGTATTTTGTTTTCCGT 227448

Qy        878 CGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTAT 937
                || || ||||     ||          |    |||       || || | | || |||
Db     227449 TTAAAAT-AAATTTGAACATTTTTTTTTTTCTTTATTTTTTATTTTGTTTGTATTTAA 227507

Qy        938 TATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAG 997
```

```
Db     227508 ATGGAAAGTAAAAAAAAATTTTATATTTAATAATTTTTACCAATTTTTTATAGTATATAA 227567

Qy        998 ATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCA 1057

Db     227568 AATATGACATTTGTTTTTATTTATTTTTATTTAATTATATTTTATTCTTCTATTTAATAT 227627

Qy       1058 TGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCA 1117

Db     227628 TTTTTGTTTGTTGTTTGTTTGTTTTATTTTTTACAGTTATATTTATATATTATATATACA 227687

Qy       1118 TATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACT 1177

Db     227688 TAATATTTGTATATAAAAAAATATAAAGTGTAATTTTTTTTTTAATTTTTTTTTTGTTAAC 227747

Qy       1178 TTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219

Db     227748 ATTGCTTAATAAATAGTATTCTAATAAATAAATAAATAAATA 227789


RESULT 10
AF324424/c
LOCUS       AF324424               3026 bp    DNA     linear   INV 27-FEB-2002
DEFINITION  Ichthyophthirius multifiliis immobilization antigen isoform
            (IAG52A) gene, complete cds.
ACCESSION   AF324424
VERSION     AF324424.1  GI:12698726
KEYWORDS    .
SOURCE      Ichthyophthirius multifiliis
  ORGANISM  Ichthyophthirius multifiliis
            Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;
            Hymenostomatida; Ophryoglenina; Ichthyophthirius.
REFERENCE   1  (bases 1 to 3026)
  AUTHORS   Lin,Y., Lin,T.L., Wang,C.C., Wang,X., Stieger,K., Klopfleisch,R.
            and Clark,T.G.
  TITLE     Variation in primary sequence and tandem repeat copy number among
            i-antigens of Ichthyophthirius multifiliis
  JOURNAL   Mol. Biochem. Parasitol. 120 (1), 93-106 (2002)
  MEDLINE   21839613
   PUBMED   11849709
REFERENCE   2  (bases 1 to 3026)
  AUTHORS   Lin,Y., Lin,T.-L. and Clark,T.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2000) Microbiology and Immunology, NYSCVM Cornell
            University, Ithaca, NY 14853, USA
FEATURES             Location/Qualifiers
     source          1. .3026
                     /organism="Ichthyophthirius multifiliis"
                     /macronuclear
                     /mol_type="genomic DNA"
                     /serotype="D"
                     /isolate="G5"
                     /db_xref="taxon:5932"
     gene            1285. .2817
                     /gene="IAG52A"
     mRNA            1285. .2817
                     /gene="IAG52A"
                     /product="immobilization antigen isoform"
     CDS             1330. .2736
                     /gene="IAG52A"
                     /note="52 kDa protein"
                     /codon_start=1
                     /transl_table=6
                     /product="immobilization antigen isoform"
                     /protein_id="AAK01661.1"
                     /db_xref="GI:12698727"
                     /translation="MKNNILVILIISLFINQIKSANCPVGTETNTAGQVDDLGTPANC
                     VNCQKNFYYNNAAAFVPGASTCTPCPQKKDAGAQPNPPATANLVTQCNVKCPAGTAIA
                     GGATDYAAIITECVNCRINFYNENAPNFNAGASTCTACPVNRVGGALTAGNAATIVAQ
                     CNVACPTGTALDDGVTTDYVRSFTECVKCRLNFYYNGNNGNTPFNPGKSQCTPCPAIK
                     PANVAQATLGNDATITAQCNVACPDGTISAAGVNNWVAQNTECTNCAPNFYNNNAPNF
                     NPGNSTCLPCPANKDYGAEATAGGAATLAKQCNIACPDGTAIASGATNYVILQTECLN
                     CAANFYFDGNNFQAGSSRCKACPANKVQGAVATAGGTATLIAQCALECPAGTVLTDGT
                     TSTYKQAASECVKCAANFYTTKQTDWVAGIDTCTSCNKKLTSGAEANLPESAKKNIQC
                     DFANFLSISLLLISYYLL"
ORIGIN

  Query Match             4.7%;  Score 57.8;  DB 3;  Length 3026;
  Best Local Similarity   44.0%;  Pred. No. 1.6;
  Matches  341;  Conservative    0;  Mismatches  427;  Indels    7;  Gaps    2;

Qy        445 TTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTGGTCACTCTATGAACCACAATCGTA 504

Db       1325 TTTTTAAATATATATAAATAATAATTTATGGAATTTTATATGAAATAAATATAATTATTA 1266

Qy        505 ACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTCCTCTTTGTGGTGGTTTCTCTAATA 564

Db       1265 ATAAAATTTTATTTTTAATAACTCTAAAATAAATTTATTTAGATATTAGTTATCTTAATG 1206

Qy        565 AATGGTGGGGTAGGAAGCACAATCAACATCATATGTTCACAAACAACATTCTAAAGGACG 624

Db       1205 TTTAGATTTTTAATAAATAAAATATTATTATTTATATTGTATTACATATGCTTTTTTTAA 1146

Qy        625 AAGATATCTAACACGATTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAAT 684

Db       1145 AAAAAAATTAATCATACAATAATTTTTATGAAATTAATAATTAATTTTTTATTTATATTT 1086

Qy        685 TAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGG 744

Db       1085 TTATAAATATTTTTTCTTCTAAAAATTTATAATAATTTTTATATTGTACTTATTACTTAT 1026

Qy        745 TATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCT 804

Db       1025 ATTATTTCTTTTAATAAAAATAAGATATCATAAAATTATGAATTTTAAATTTTTTTTTGT 966

Qy        805 TCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCA 864

Db        965 ATAATGATTTTTATATTATCCTAAATAATTAATATACAATATTATTTTAATTATTTTATT 906

Qy        865 CTTTACCATTT-TTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATA 923

Db        905 TTTTATAAATTATATTCACATTTTAAAAATTATAGTGATGTCAACATTATGACTGATTAA 846

Qy        924 TTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAA 983

Db        845 TATTACATATTCTTCTGTATTTTCAAAAATAAATAGTCAGAATAAAAAAATTAGTTATAT 786

Qy        984 ATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGG 1043

Db        785 AGCACAAACATCAATATAATAATTTTGTTATTTATTTCAAATTAAAAAATTATATATTGA 726

Qy       1044 AACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAAT 1103

Db        725 AAGATTTATTCTTTTGAA------ATTTAAATAAATTATTTATATGGAATAGCAATTAAA 672

Qy       1104 TTATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTT 1163

Db        671 TATGTAAAAATAATAAAATATAAATTTCTAATGTATAAAAAACAAAATATTATATAATTA 612

Qy       1164 GTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAA 1218

Db        611 ATTAAATAATAAATTACAAATAATATTTAATATTTTATTTAAATAAAATTAAAAA 557


RESULT 11
AX598900/c
LOCUS       AX598900               8056 bp    DNA     linear   PAT 14-FEB-2003
DEFINITION  Sequence 240 from Patent WO02077272.
ACCESSION   AX598900
VERSION     AX598900.1  GI:28399038
KEYWORDS    .
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1
  AUTHORS   Berlin,K., Braun,A., Distler,J., Guetig,D., Howe,A., Mueller,J.,
```

```
            Olek,A., Piepenbrock,C., Adorjan,P., Grabs,G., Lesche,R., Leu,E.,
            Lewin,A., Lipscher,E., Maier,S., Model,F., Mueller,V., Otto,T.,
            Pelet,C. and Ziebarth,H.
  TITLE     Methods and nucleic acids for the analysis of hematopoietic cell
            proliferative disorders
  JOURNAL   Patent: WO 02077272-A 240 03-OCT-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .8056
                     /organism="synthetic construct"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
ORIGIN

  Query Match             4.7%;  Score 57.4;  DB 6;  Length 8056;
  Best Local Similarity   44.3%;  Pred. No. 1.6;
  Matches  283;  Conservative    0;  Mismatches  351;  Indels    5;  Gaps    1;

Qy        586 ATCAACATCATATGTTCACAAACAACATTCTAAAGGACGAAGATATCTAACACGATTACA 645
              || || |   ||| |    ||  |   || | ||  |  || |    |||    |  | |
Db       2746 ATTAATAATTTATTTAACGAATTATTTTTATTAAAAAAAAAAAATAATAAATTTAAAAAA 2687

Qy        646 AATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTT 705
              | ||       |     || | |  ||| || |   ||      |   | ||   || ||
Db       2686 ATTTAAAAATTTATTATTTTTAAAATTTTAAATTTTAACGTTTTTTTTTATTTTTTTTT 2627

Qy        706 ATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACT 765
              | |   ||||  |   ||  |  ||   ||    |   | ||||| || ||  |
Db       2626 TTAAATTATTTATATTTATAATTTTTTTTTTCGTTTATTTTTTATTTTTTAAAAATTTTA 2567

Qy        766 TCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTG 825
              | || ||  | |   ||| | | ||  ||  |  ||| |   |  || | |  || | |
Db       2566 TTTAAATTTTTAAATTTTTTAATTTATTTTATAATTTAATTTATTTTTTTTTATTTTTCG 2507

Qy        826 GAAA-----TCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGA 880
                ||     | || |   | ||| |||||   ||| || |   |  | |  ||      |
Db       2506 TTAATATTTTAATTATTTTAAAACGAAATAATAAAAAATATATAAATAAAAATAAAAAAA 2447

Qy        881 ACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTAT 940
              |  | |  |       |   | ||| ||   |||     |  |         || | |   |
Db       2446 AATTAATTTTTAAAAATAAATAAATTATTTTTTTTTTAATTTTCGAAAAATAATAAAATT 2387

Qy        941 GGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATT 1000
              |    | |     | |   || || |  |||    || || |    | ||   |
Db       2386 TAATAAATTATTTATAAAAATAAAAAATTATATTTAAAAATTAAATAAAATTTATTTAAA 2327

Qy       1001 ACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGA 1060
                | |||  |  |    || ||    || || ||||  |    |      |   ||
Db       2326 TACGAAAATTAATAAATTATTTTAAAAATAAATAAATTAAAAAATTAAAATTTTTTTAAT 2267

Qy       1061 AGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATAT 1120
                ||  | | |||||    |  |||   || |  || ||| ||||   |  || |
Db       2266 TTATTTTAAATATTAAAAAATAAATTTTTATTAAAATTAAATTTTTTTTTTTTTTTATTT 2207

Qy       1121 TTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTT 1180
              ||| ||      |  | | | || ||  |||  | | ||  ||| |||  ||||| || |
Db       2206 TTTTTTTCGTTTTTTTAAATTTTTAATAAAATTATAATTTTTTTTTTTTTTATTTTTTAT 2147

Qy       1181 TAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAAA 1219
              |  |   | || | |   |||  ||||  | |  ||| |
Db       2146 TTTTAATAAAATAAAAAATTATTAAAATTATATTAAATA 2108


RESULT 12
AC116979/c
LOCUS       AC116979               331039 bp    DNA     linear   INV 21-FEB-2004
DEFINITION  Dictyostelium discoideum chromosome 2 map 6445720-6776760 strain
            AX4, complete sequence.
ACCESSION   AC116979
VERSION     AC116979.3  GI:42733765
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 331039)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   5  (bases 1 to 331039)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2004) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On Feb 21, 2004 this sequence version replaced gi:28829703.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES             Location/Qualifiers
     source          1. .331039
                     /organism="Dictyostelium discoideum"
                     /mol_type="genomic DNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /chromosome="2"
                     /map="6445720-6776760"
     CDS             complement(join(912. .1165,1280. .1427))
                     /note="GeneID exon scores (in order of location ranges):
                     30.46, 12.19 - GSCJ_ID dd_02783"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAS38697.1"
                     /db_xref="GI:42733766"
                     /translation="MSEKTFTRQEVAKHCSLNSLWIIYNDDVFDVTNFVVEHPGGEEV
                     LKGNGGKDATQEFDDVGHSASAIAKMQSLRIGRIEGAKPREEKKKEIKKTTTTTSAPK
                     QQESAGLGLLKIPLIIIVLAIAAYFFMGDQQ"
     CDS             complement(join(2345. .3098,3189. .3526))
                     /note="GeneID exon scores (in order of location ranges):
                     98.40, 31.82 - GSCJ_ID dd_02784"
                     /codon_start=1
                     /product="similar to Dictyostelium discoideum (Slime
                     mold). Protein disulfide isomerase precursor"
                     /protein_id="AAO52220.2"
                     /db_xref="GI:42733767"
                     /translation="MKILLFVTLIALAFVALCSAEGNVVVLSPDNFDTVVDGSKTVFV
                     KFYAPWCGHCKKLAPDFEILADTFAPVSNKVVIAKVDCDQADNKALCSKYDVSGYPTL
                     KIFDKSTTAKDYNGARSVDELLTYINNHAKTNVKVKKAPSNVVDLSPSNFDSVVLDKS
                     KNVLVEFYAPWCGHCKKLMPDYEILGNTYANEKDVVIAKIDCDAADNKAICSKYGVTG
                     FPTLKWFGKQSKDGEKYEQGRDLDTFINYINKQAGVNRVKGGKLAVGAGRVEQLDTIA
                     TEFIAAAAEVRKELVKKAQTVVDSLPEELRTEGSYYVKVMKTIAEKSIDFVTTEIARI
```

```
                TKLVSGSMSGKKADEFAKKLNILESFKSK"
CDS             join(5795. .6179,6292. .6567,6659. .6881,7332. .7363,
                7523. .8215,8284. .10451)
                /note="GeneID exon scores (in order of location ranges):
                14.11, 5.65, 10.80, 0.73, 45.28, 141.12 - GSCJ_ID
                dd_02785"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Hypothetical 127.0 kDa protein"
                /protein_id="AAO52221.2"
                /db_xref="GI:42733768"
                /translation="MNINLLKFFIIYYFCFVLINFCNCQTLKITDISNPKNYKVYPQY
                SDREFCDFSFEILCYVEPAMDEGDQLMFTTDFDYSEMIRSDRTGSIIGITHTALSFGS
                YTSNIVCYLLNNPTIKVDLNISYSCEMIDYELMKVKILKPNYYFKKNNLYGNGIIQLIG
                LKYPLNNLVIFGGSAYPLNPNYFLISLEELDGGTVENYNISLLFYGPHTIVVTIPAST
                YLDYNLDISNSKINYYPNETTNVSEFGLGCSTVYSVTINKTSNDYEPYLYVDDGNSKN
                ADFTPIYQDENFITYLGQFNSFVTVTSPYAQILNSSILSTLLEFECKLLFDESYLFRF
                KVNTNGYGASIYLGSDITLRYESVVDSESNSSVLEFETVVDLYKDSITKFTISNLLSN
                RINYLRGEYYSISPLLKFEIPFGELNPVLITNISFLYNNIDVTNKSVNNIIYFSYEGE
                FDKGSSFALLLNDPVSILKNGDLDSSDFRFSKWNSSTSMFQIEFTVPANTQPGHIPFK
                LYGNSNPNYNPIDSTMLPLSAQLFIISSNFDSYGPIFNKVEKINTTNEFGWKFVIGDK
                INGFDYGDIIVRGEMDSSTYKFHLTPLNLTSGDKYNGEYQINIKISILCASQNYIITD
                VYLFDTRGNFAKFSVSEQCESIRNPFINFLGDSTINKIYKKCTTENDGYDISPPELTL
                FSAVKTILNDSQTIVIAFDFQAVDLDTGLKNEQYPIVYIESTELQIIQCESKIKSINS
                KTANYTCEIELPVGFGYNSDIVFSVYGFINNGGYYSGFPSTYISSAFPIYSMSNIQLV
                RKIEILRTSQISSTGGELWIVGRGFKTSIQYVQIKYSTDLVFSQISFPSIIFSTALLI
                LDIDPTDKPFTIRVVTKNGQQSNEFIVTPLFSFLSVSPLNTPTPTTTTQLPTPTNKPQ
                TCLGEPICGGPKQGFCSSKGCVCYQPWIGNDCNAQVVIIPQPSTNTSQPSTEIPIIDN
                NNNQQPSGINYLYKSLISIVSIRELDFNGNQINLYPFDRWIFTQINQNKSQYFTSIKN
                NSTTPFNTNITVILEWFKDTSTIEFANSSITMYPSSIKYTIEITEYKFSNQLNNLQLV
                MSASFESSKTKEICSLKDFGETSNGDNSNYFKIQINDHSLYGRFIKRAIIDSKISTIE
                NVLLDSKMNSIQTASSSQSYIGITIPNYQKSIIIDPDFSVLIDSKSTSKNDENSVCTP
                NQSKLTTPQLIGIIIGSIAFAAVIAISIAYHIIKKKKDTKFKNGLQNKLKTFKS"
CDS             join(11030. .11042,11114. .11283)
                /note="GeneID exon scores (in order of location ranges):
                -1.76, 12.26 - GSCJ_ID dd_02786"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAO52222.2"
                /db_xref="GI:42733769"
                /translation="MTILKSISSMGISNRKNSFSFVSTNSMIETNQNVNSVSKSGYSK
                LIKGAALMAEGISGFF"
CDS             complement(11515. .12534)
                /note="GeneID exon scores (in order of location ranges):
                87.29 - GSCJ_ID dd_02787"
                /codon_start=1
                /product="similar to Burkholderia sp. DBT1. Putative
                oxidoreductase"
                /protein_id="AAM34303.3"
                /db_xref="GI:42733770"
                /translation="MLPITNKVALITKKGFSESFSIVDNEIPKPKGNEVLIKVDYVGL
                NRCDLYFSLGLYLPFENYPNVSGFEASGTVVEFANDELENSSKFKVGDKVSVIYTCDH
                RKYGTFAEYAVFPIGCLVKNNELSSQKEIAANWISFFTAYYGLVEFAQIKKGDYVIIT
                AGSAGTSLAGIAIAKFFGAHTIVTSRNEKKKQQILDYGADYFVSLDTEDIVKRTFEIT
                NGKGANIIYDSVGGQSAQKLFESSSQFGKIVFYGNMNHSEPTPLPILIGLKNYLTIKV
                LNTNEYLYQPEFNKKAIEFLNSNFHHFKSIVGKEFIGIDSLVDALKYLETGDLFGKVV
                VKVNN"
CDS             complement(join(13231. .14884,15364. .15368))
                /note="GeneID exon scores (in order of location ranges):
                104.49, -7.13 - GSCJ_ID dd_02788"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAM34304.3"
                /db_xref="GI:42733771"
                /translation="MINPNLLKPLFIENFCSSSQSWLDLIKNRKLNFYTITCLTHYYS
                TVKVLYFDIKSIFENDKKQNENQQNEKQQQQQQQQQQQQQQNEKSLNDRKSIFKKSYH
                IIRLLIEMNDIINYKDIRIYLPNQSEERKLLMKIREGEEIDIQSCFELVQSKYLEFES
                KFNELKKNSTNDSLIKLNSARALHVLNDWVVDIRRNSIENSNNLNSFKKYKFENENHH
                DESNNSFKIAMDLMNKNNIKGTLIYFGPSGSHLHGLQDDDHNDENSNDNEKIDWIGIY
                AAPTDDIVSLFPLALKHDSSGFIMQSNEHIGINLNPAPDRPNNQNKKDSFSKGIILYE
                LKYALQLIVNCSYRVLECVFMSSNQVLNNHPFKNKFYITDAWKDFINQVLINDNNNNK
                TFSKLSNSNLIQQYWGLAKGISMKNSTIFKMVKSNLLRKKDDAITIAEEDLVTDDSIS
                KNLFFLNYLLIQCNTILNGIVFNENNNTIINSPNNEILLKEKEIPSIFIDKQSKDYQFL
```

```
                MSLKHNKLNKLSIIQSLIIDSDSILEEIRIKLDTLKKIQRIDLV"
CDS             join(17515. .17664,18062. .19243)
                /note="GeneID exon scores (in order of location ranges):
                1.18, 93.30 - GSCJ_ID dd_00463"
                /codon_start=1
                /product="similar to mitochondrial carrier protein family;
                protein id: At1g14560.1 [Arabidopsis thaliana]"
                /protein_id="AAO52223.2"
                /db_xref="GI:42733772"
                /translation="MNSSDFKKSFKESTENNSNTYRPSKTLNTYKSLKNGNNIANSLI
                IQLNNNYLVSGSIAGAISRSATAGFERLTIIQQVQGMSQNLSQGYVGCIAAMKEMVKR
                EGFKSIWKGNGANIVKVSPNSGIRFLTYEFCKKHFLDNSSNHPSSSSIENGIDGNGVG
                CGSGSEMKMTVPQTMFSGAMAGLTSTFFTYPLDVVRIRLSLQGSCSNDYAAHRYNGIT
                HSFFKIHKDEGVKGLYKGLGTSIASIVPWVSISFATYEGFKIICKKMILNYQISSSSL
                STTTTTPSSINNNNNNNNKNNNSFIYENELGENGINLTNTSGCSTMASTMPSSLLINS
                VASDENELKKGVNMICDFVCGALSGAVTMTVCYPLDVLRRRMMIQGIGGNKVLYKNGW
                DATKKILSNEGLVAFYHGIIPAYFKVVPTVAISFAVYEICKDLGSNKYQQK"
CDS             join(20299. .20406,20551. .20660,20960. .21569,21708. .21828,
                22984. .24188)
                /note="GeneID exon scores (in order of location ranges):
                9.06, -2.15, 55.50, 6.70, 58.02 - GSCJ_ID dd_01980"
                /pseudo
                /codon_start=1
CDS             join(24256. .24266,24388. .24538,24623. .24750,24902. .27884)
                /note="GeneID exon scores (in order of location ranges):
                -2.62, 6.90, 11.20, 259.32 - GSCJ_ID dd_01978"
                /codon_start=1
                /product="similar to Dictyostelium discoideum (Slime
                mold). Histidine kinase DhkF"
                /protein_id="AAO52224.2"
                /db_xref="GI:42733773"
                /translation="MILKFIIQQCVFKVAKIFSDNLIQIFMDISSVYSSSTLFFSKLS
                LLMCKVFNSKCAGISVLNKDGLFDTIAFCSSTQELENFSYSLKKISKNITYYYDYMEIP
                LKDSKEFYFGDIVLFGDQTLNPNILDTNILTFITNRTSMEFEKKKVSDQLIIAKNLLD
                QSPSCSCLVSRNGKIIRKIGAFYERLLGLEIGENISSVEKSSDISTILKQVCFENNNL
                STETTITKKNGEKYPAEVFVKEISDIHSNSIGIMIIVRDITDQIRLKEMNIELQKKSE
                LEQKRNQELMEARDLALTATKIKSQFLATISHEIRTPLNGIITMGEMLLSTSPLNTEQ
                HDIAETIFGSSELLLSITSDILDFSKIEASKLELEMIEFDFIGCLEGIGKTIGVSITN
                KPIEIAFLMDTDIPHRLIGDPNRLIQIMLNMGTNAVKYTDRGHIVFRISVISREQNRC
                KIKISIDDSGIGIGEDQRAHLFEPFHQIDSSSTRKYGGSGLGLAISSKLAKLMGGEVI
                LERSKPGVGSLFSVTLNFEQIGSNTLKSLLPDKSFSRNKTCIILDNYEYTASIASQRF
                DQIFEESNIILVKESTVKKFFELVQEYKINGDVNKIDLKNLDPELLPFFDSSLICIIV
                FHRFMDNLDIFLKYVKDFIEFYKKKIVVALGINHKNFKNLPKRRDFLIFKKPISSTNL
                IKVSNMARKIIPKSSSSSNLIQTISQIDNQQQQQQQQLQQQEQEQQHQQQQLQQEQQF
                VTISPHSDSSEKKTTPKKDRGKYDFNISPLRIDRFGTESTSSPKIKLYSDTSSDSGES
                DEFEFSENLRELKSDGELVVLKNQPYKEQRLSNIQLINDPIVTPPTPKANSEVVSTKA
                ESTFFQSSSITIKETRSDSVNSVNSVNSVNSVNSLNSENSNNSLDVGIRTRPRILLVD
```

Query Match 4.7%; Score 57; DB 3; Length 331039;
Best Local Similarity 43.8%; Pred. No. 0.86;
Matches 249; Conservative 0; Mismatches 320; Indels 0; Gaps 0;

```
Qy        641 TTACAAATTGTGGTAATTCCCCTTCTTATTTTTAAAGTGGAAATTAGACTCCATCTTAGC 700
              ||  || ||||  ||| ||    | || ||||      |    | ||    |   | |||
Db     250435 TTCAAATTTGTATTAAATCTATCTATTTTTTTATTTTTTTTATTTTTTATTTTTTTTATT 250376

Qy        701 TTCTTATTATGAATTTGAAGGAATCTTCCTTGCCTTGCACTGGGTATTATTATTCAACTA 760
              || |||||     ||  ||  | | ||  || | |  | | | | ||||   |||  | |
Db     250375 TTTTTATTTATTTTTGAAATAATTGTTAATTACGATTAAGTTGTTATTTCAATTTCATTT 250316

Qy        761 AAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTTCAGTGCTTCTATTCT 820
                |    |||||   | || | | |   |  |||  |   || || | |  ||   || |
Db     250315 CTATAAATATATTTAATTTTTATTTTTTTAAATTTTTTTATTTTTTATTTTTTTATTTAT 250256

Qy        821 TGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCACTTTACCATTTTTCGA 880
              | ||  ||| |  ||  | || |  ||    |      | |   | |   ||||||
Db     250255 TTTTTAAAAAAAAAAAAGCTTAATTATAACAAAATTTTTTATTTTTTATTTTATTTTTTTT 250196

Qy        881 ACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATTCTCTCTACTTATTAT 940
              | ||| |    ||   || || |    |||  |      | || | | |  || || |
Db     250195 TTTTAAAAAAATATCATTAAGTAATTTTTTTTTTTTAATTTTTTTTTTTTTTTTTTTTT 250136

Qy        941 GGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAATTCCTTTCTACAGATT 1000
```

```
Db     250135 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATTAACGACCACCTT 250076

Qy       1001 ACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAACCTTAAGATTCATGA 1060

Db     250075 TTACTAATCCCAACAATTCTTATAGATATTTAATAGGTTAAAAAAAACAAACAAATTTGT 250016

Qy       1061 AGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTTATTTTAAATGCATAT 1120

Db     250015 ATTATTAAATAAAATAAAAAAAAAAAAAACCATTTTAAAAAAAAAAAAGATAGAATTTTAT 249956

Qy       1121 TTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGTTTACTTATTACTTTT 1180

Db     249955 TTTATTTTTTTTTTTAAAAAATCAAGAAAATGTCAGATGATAAAATTCATTTAAAAGTTA 249896

Qy       1181 TAATCTGAGAAAACAGTCTTAACAAAAAA 1209

Db     249895 GATTGGTTGATAAGAAAGTTTTTAAAAAA 249867
```

RESULT 13
AF362375/c
LOCUS AF362375 7079 bp DNA linear INV 19-APR-2002
DEFINITION Dictyostelium discoideum histidine kinase DhkE (dhkE) gene, complete cds.
ACCESSION AF362375
VERSION AF362375.2 GI:20198941
KEYWORDS .
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE 1 (bases 1 to 7079)
AUTHORS Anjard,C. and Loomis,W.F.
TITLE The histidine kinases of Dictyostelium
JOURNAL (in) Inouye,M. and Dutta,R. (Eds.); HISTIDINE KINASES IN SIGNAL TRASNDUCTION; Academic press, San Diego (2001) In press
REFERENCE 2 (bases 1 to 7079)
AUTHORS Anjard,C. and Loomis,W.F.
TITLE Direct Submission
JOURNAL Submitted (19-MAR-2001) Biology 0368, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA
COMMENT On Apr 19, 2002 this sequence version replaced gi:14133661.
FEATURES Location/Qualifiers
source 1. .7079
/organism="Dictyostelium discoideum"
/mol_type="genomic DNA"
/strain="Ax4"
/db_xref="taxon:44689"
gene <765. .>5992
/gene="dhkE"
mRNA join(<765. .1637,1766. .>5992)
/gene="dhkE"
/product="histidine kinase DhkE"
CDS join(765. .1637,1766. .5992)
/gene="dhkE"
/codon_start=1
/product="histidine kinase DhkE"
/protein_id="AAK54094.2"
/db_xref="GI:20198942"
/translation="MDKLKINNNLSPPSSPSSSTTTPNLSSTNLENNLNSNINNNINN
FNLNNSTNTFNNSNNIIINNNNNNNNNNNNNNNNNNNNSNNNNNINNNNPNVNSPNE
VILNNNFLKQNKKGFFKRNKKMIEIFLPPKFKSPIYRRCCILLGECGVYVFIYMFFLI
FLRSFYPIFICGIVSMIVLYIVSTKTTKYKLVALIYIFVQSILNFTFFLQITNNFNNN
TVNINNTDDYSNEVIEFSKLNFLFIMNLILSLISIQIFFPKFTFSITLTCSLNIFNII
IHLISIYSLNTKKLIVFQNLIVPITVSFLLSFYSYILSIDNQEIEAKEKRFRNIFDTS
SEALVIHKNGLIIDVNSTFEKIFQIKQSDLINGTIWEYLPELESFFQNTNGNSKYTNQ
SLQQQLQQQQQQQQQLYNTITNGGNNKQTSTTSANSTPRYNNYNNNNINSNNLNNLNV
FNNLNNLNNLNNNNNNHHIHKVSRSNTGLSDSSSIFDDSDSVDFFSMKNINKCPIVLD
TVGITPCGSEFCAQVKIERKKDMLIGSGTSSFNLLNNNNNNNNNNNNNNNNNNNNNN
NNNSNNNNNNNSNNNNNNNNINNNISNNIINNNNNNNNSGIITNLSQQSNNLSTTNLN
TSSKIINHKHKNSKQDFDVMSIVDISAKKKLMIADQALRRTEELNQAKINFLTTVSHE
VRTPINGILASVEILDGSQLDVTQRDFLSCIKQSADYLLDLITDILDFSKIEAGKFEL
DRVEFNLITMLEESINIVYRTAQERGIEVLTFIDPDVPIILIGDPYRVKQCVLNFLSN
AIKFPHKGQVMVKVTIVDPINNNNNNNNNNNNNNNNNNNNNNNNNNNNKIVGFSNN
NNNNNNNNNNGNNGILNFKLSFSVEDSGVGIKEEVLDCLFTPFHQLGGSPRKYLGTGL
GLSISKKLTTLMGGEIGVKSVYGVGSAFSFTSILSTTSTTPISLQSLGSSLSIALNNL
SPTIPKISGFIYDDNIHTSNSLFNFLKLMNINLKIINPIINNNNNNNNNNNNNNNNNN
NNNNNNNNNNDNNNNNNNNDNNIDQQQLKFEFENEIDKYCNVEDDNNNVILIITNQ
LSNDNLNYFKDKINQLNNSIYWFVLCDNGLKAIDPFYYGIVKKPNNLLNLVDTVFKVY
NCQLPQDLYQLLTNGSKDDYKNYLKYRNVVIKKITEDNEKKQQQQQQQQQQMGDTLSS
TKSPQYTNLPPLDISSSSNGSLNKSNRSNLLRKSSSVYSDHVAITRGMVQVNRSPRPS
TPPPLFNLKGNNSNPNSTELNSTNSVNGNPNNDSTLETIEPLNISETVGDLCSSGSVI
NNSNFIINNNFLTSSPYGSCLTPNSGTTSPSIPININNNDNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNSFKKLEPDNFSPPPSILINDNNIIPALNVTTPITISTSTSSSSS
LLINNEINNKSTTSKRPSFIPPLDIQSKKLIINTADENHKNNCIIDFEQIFGQDGKQQ
PQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQKLNIGNQQQQQILSSPRL
QSQLLSNCDSLGNTPPVTPHRRNALIVDDTELNRKVLAQLLRRMDWSVSFAENGIEAL
KEITSERCFDIIFMDCQMPILDGFETTKLLRLRELENNWKPLNIVALSADSSSSFGQV
CFDCGMNGYLGKPITLITLKDTLLKWGGYRD"
misc_feature 2873. .3736
/gene="dhkE"
/note="Region: catalytic domain"
misc_feature 5615. .5971
/gene="dhkE"
/note="Region: receiver domain"
ORIGIN

Query Match 4.7%; Score 56.8; DB 3; Length 7079;
Best Local Similarity 47.0%; Pred. No. 2;
Matches 207; Conservative 0; Mismatches 232; Indels 1; Gaps 1;

```
Qy        746 ATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTT 805

Db        791 ATTATTATTAATTTTTAATTTATCCATATTAATTCTATTTTTTTAAATTTATAGGTTATT 732

Qy        806 CAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCAC 865

Db        731 TTATAAATTTACTTTCTCTTTAAATAATAAATAGAAAAAAAATTGAGGAAAAAAAAAAAA 672

Qy        866 TTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATT 925

Db        671 AAAAAAAAAAAGGACCCACCCATCAAAAGAAAGTGATTGGTATTTTTTTTATTTTTTTTT 612

Qy        926 CTCTCTA-CTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAA 984

Db        611 CTTCAAATATTATTTAATTTCCCTTTTTTTTTTTTTTAATTGAAAAATATTTATCATAAA 552

Qy        985 TTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGA 1044

Db        551 TAATTTAAAATAAAACCAAATCAGATCAAATCAAAAAAAAAAAAAAAAATAAAATTGTGTA 492

Qy       1045 ACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATT 1104

Db        491 TGCCAAATATTTTTAAAAAATTGTTATTTTTACTTTCACTAATGTATTAATGTTTATTAT 432

Qy       1105 TATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTG 1164

Db        431 TATTATAATTATTATTTTTATTATTATTTTTATTATTTATTTATTTTTATTATTATTATT 372

Qy       1165 TTTACTTATTACTTTTTAAT 1184

Db        371 ATTATTATTATTATTTTTAT 352
```

RESULT 14
CR387984
LOCUS CR387984 159835 bp DNA linear HTG 19-APR-2004
DEFINITION Danio rerio clone DKEY-230E6, *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.
ACCESSION CR387984
VERSION CR387984.3 GI:46408734
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 159835)
  AUTHORS   McLay,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-2004) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 19, 2004 this sequence version replaced gi:46358283.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zK230E6
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 157552 bases at least Q40
            Consensus quality: 157753 bases at least Q30
            Consensus quality: 157954 bases at least Q20
            Insert size: 159335; sum-of-contigs
            Insert size: 161212; 5.4% error; agarose-fp
            Quality coverage: 8.49x in Q20 bases; sum-of-contigs Quality
            coverage: 8.53x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    83704: contig of 83704 bp in length
            *    83705    83804: gap of 100 bp
            *    83805    90391: contig of 6587 bp in length
            *    90392    90491: gap of 100 bp
            *    90492    95482: contig of 4991 bp in length
            *    95483    95582: gap of 100 bp
            *    95583   120943: contig of 25361 bp in length
            *   120944   121043: gap of 100 bp
            *   121044   151145: contig of 30102 bp in length
            *   151146   151245: gap of 100 bp
            *   151246   159835: contig of 8590 bp in length.
FEATURES             Location/Qualifiers
     source          1. .159835
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-230E6"
                     /clone_lib="DanioKey"
     misc_feature    1. .83704
                     /note="assembly_fragment:01147
                     fragment_chain:1"
     misc_feature    83805. .90391
                     /note="assembly_fragment:00125
                     fragment_chain:1"
     misc_feature    90492. .95482
                     /note="assembly_fragment:00063
                     fragment_chain:1"
     misc_feature    95583. .120943
                     /note="assembly_fragment:00330
                     fragment_chain:1"
     misc_feature    121044. .151145
                     /note="assembly_fragment:00682
                     fragment_chain:1"
     misc_feature    151246. .159835
                     /note="assembly_fragment:00215.0"
ORIGIN

  Query Match             4.7%;  Score 56.8;  DB 2;  Length 159835;
```

```
  Best Local Similarity   42.6%;  Pred. No. 1.1;
  Matches  202;  Conservative    0;  Mismatches  272;  Indels    0;  Gaps    0;

Qy        746 ATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTGAATTGATTGCTGGTTTCTT 805
              |||| ||   || |||||      |     ||| ||  |   |   ||  |   |  ||
Db        376 ATTAATAAAAAAATAAAAANNNNNAAAANATATTTTTATTTTTTATTTTATTTAATAATT 435

Qy        806 CAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGAAATTCGAAAGAAGAATCAC 865
              | |  || | ||        || | ||   |   | ||  |  | | | |  ||
Db        436 TATTTTTTTTTTAAAAAATTAATTTATTNTATANATATATATATATANATATATATATA 495

Qy        866 TTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACTACGCTTTCCACGACATATT 925
              |  | |     || ||  |          |  | |||  |   ||   |    |  |
Db        496 AAAATAAATAAATAAATATNNATATATANANANTAAAAAAAAAATTTNNNAANNTANTTA 555

Qy        926 CTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAACATCACTTTTTCCCATAAAT 985
              | | |   | |  |   | | | |   ||||  | |  ||    |  |      |||  |
Db        556 TTTTTTTTNTNTNTTTNTTTTTTTATTTAATTATATTAAAATATATATANAAATATATTT 615

Qy        986 TCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTGAAGAATTAAAGAAGTGGAA 1045
              | | |  | |  |   ||   | | | |  |    | || ||| ||  ||    ||
Db        616 ATTTNTATTAAAAAAATAAAANANTTTTATAAATAAAANTNAAAAATAAATAAATAAAAA 675

Qy       1046 CCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACCTTTGAAAATAAATAAATTT 1105
              |  | ||  ||  |        | |  | | || |    |||  |  |   ||  |||
Db        676 AATATATATAAATATATATATATATATATATATTTATTTTTTTTATTTTTTTATTTTT 735

Qy       1106 ATTTTAAATGCATATTTTATTAGTAATACTAACAATTGTAGGAAATGTGTTATGGTTTGT 1165
              ||||   |  |||| | |  |  || |  || ||    |  |||| | |  |    | |
Db        736 TTTTTTTTTTATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATATATTTACATAT 795

Qy       1166 TTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAAAAAAAAAAAAAAAAA 1219
              ||  |||   | | ||     |  ||  |     || ||||||||||||||||
Db        796 ATATATATATATATATATAAAAAATAATAAAAAAAAAAAAAAAAAAAAAAAAA 849


RESULT 15
MBREV/c
LOCUS       MBREV                   76568 bp    DNA     circular INV 14-FEB-2003
DEFINITION  Monosiga brevicollis mitochondrion, complete genome.
ACCESSION   AF538053 AF275274
VERSION     AF538053.1  GI:23344058
KEYWORDS    .
SOURCE      mitochondrion Monosiga brevicollis
  ORGANISM  Monosiga brevicollis
            Eukaryota; Choanoflagellida; Codonosigidae; Monosiga.
REFERENCE   1  (bases 64800 to 65249)
  AUTHORS   Bullerwell,C.E., Burger,G. and Lang,B.F.
  TITLE     A novel motif for identifying rps3 homologs in fungal mitochondrial
            genomes
  JOURNAL   Trends Biochem. Sci. 25 (8), 363-365 (2000)
  MEDLINE   20377911
   PUBMED   10916154
REFERENCE   2  (bases 1 to 76568)
  AUTHORS   Lang,B.F., O'Kelly,C., Nerad,T., Gray,M.W. and Burger,G.
  TITLE     The closest unicellular relatives of animals
  JOURNAL   Curr. Biol. 12 (20), 1773-1778 (2002)
  MEDLINE   22288938
   PUBMED   12401173
REFERENCE   3  (bases 64800 to 65249)
  AUTHORS   Lang,F.B. and Bullerwell,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUN-2000) Biochimie, Universite de Montreal, 2900
            Boul. Edouard Montpetit, Montreal, QUE H3C 3J7, Canada
REFERENCE   4  (bases 1 to 76568)
  AUTHORS   Burger,G. and Lang,B.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-AUG-2002) Biochimie, Universite de Montreal, 2900,
            Boul. Edouard-Montpetit, Montreal, Quebec H3T 1J4, Canada
COMMENT     On Sep 29, 2002 this sequence version replaced gi:9966505.
FEATURES             Location/Qualifiers
     source          1. .76568
                     /organism="Monosiga brevicollis"
```

```
                /organelle="mitochondrion"
                /mol_type="genomic DNA"
                /strain="ATCC 50154"
                /db_xref="ATCC:50154"
                /db_xref="taxon:81824"
gene            17. .2894
                /gene="rnl"
rRNA            17. .2894
                /gene="rnl"
                /product="large subunit ribosomal RNA"
gene            4627. .6123
                /gene="nad4"
CDS             4627. .6123
                /gene="nad4"
                /EC_number="1.6.5.3"
                /codon_start=1
                /transl_table=4
                /product="NADH dehydrogenase subunit 4L"
                /protein_id="AAN28335.1"
                /db_xref="GI:23344059"
                /translation="MWNTLIPIFALPFIGILILLILPFNNELNIKKGKQIALGTSILT
                LLESLRLWFHFDYSTSKFQYILKLNWINDELTSNSWNIVLGIDGISLVFIILTTLLIP
                ICILSSWDSVKILVKGYLISFLVIEILLIGVFTILDIFGFYILFEGVLIPMYLIIGIW
                GSREQKITAGYYFFFYTLIGSVLMLIGILYLYSISGTTDYLTLLGYEIDPTAQKYLFL
                AFFASLAVKIPKFPFHIWLPQAHVEAPVAGSVILAGILIKLGGYGFIRFTIPLLPDAS
                HYFAPLVFILGVLAVIYASLTTLRQTDLKRIIAYSSVSHMGLVSLGIFTLTANGIMGS
                IFLQLAHGLVSSALFIIVTVLYDRHHTRLVKYYRGMTVTMPIYSFLFLFFTLANIAVP
                LSCNFVGEFLCLLAIFEVNTTVAILASLGIILSACYSLFLYNRVCFGSFSTYLSNNSY
                HLLSIKDVSRREFFVLFPLVILTIILGVYPHPIFDLLNISVINLTTFL"
gene            7006. .7089
                /gene="trnL"
tRNA            7006. .7089
                /gene="trnL"
                /product="tRNA-Leu"
gene            7386. .7736
                /gene="rps14"
CDS             7386. .7736
                /gene="rps14"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein S14"
                /protein_id="AAN28336.1"
                /db_xref="GI:23344060"
                /translation="MNKIKLKFKTHSLKLDFKKRTLYVNKENNHITYNALIKNQILPL
                SVKYNLLEKGSISYEFNDKNPLSLIPTSHIRNKCLLTGRSRGNLPKYGLSRIMFTELA
                SEGLIPGMTVKNKI"
gene            8034. .8513
                /gene="rps8"
CDS             8034. .8513
                /gene="rps8"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein S8"
                /protein_id="AAN28337.1"
                /db_xref="GI:23344061"
                /translation="MNELISKINNNTNYRKMLTCQFNMDNYSILKVLKSEGVIKLLEI
                NKNNKNQKIINFLINDKLGKLGTNVFLKPISKPGRKVYISNKEILSMSNTKKGFTINT
                SKIGIKTLYSNLSEHKIYLLRTSSNNLQLLNNKGIITQEEAIKNNVGGELLLVVKFQ"
gene            10529. .11287
                /gene="atp6"
CDS             10529. .11287
                /gene="atp6"
                /EC_number="3.6.1.34"
                /codon_start=1
                /transl_table=4
                /product="ATP synthase F0 subunit 6"
                /protein_id="AAN28338.1"
                /db_xref="GI:23344062"
                /translation="MAYSAFDQFEIIRIIPLHLFGNLDISITNSTIFMFIALIYLIVL
                YSVSGIKEGSLVPSRWQSVMEIIYETVFGVVEDNIGKDKESQAYFPWILTVFMFIVLM
                NLFGIIPYTFSPTAHIAVTFGLSLSIFIGVTLLGMVNHGSNYFSMFMPAGSPLILAPF
                MIVIELISHTAKAVSLGVRLAANITAGHILFAILSGFTWTMLTSGGFMIIASLFPMAI
                VLFITIIEMAVAVIQAYVFSLLTAIYISESIHLH"
gene            11668. .11967
                /gene="atp8"
CDS             11668. .11967
                /gene="atp8"
                /EC_number="3.6.1.34"
                /codon_start=1
                /transl_table=4
                /product="ATP synthase F0 subunit 8"
                /protein_id="AAN28339.1"
                /db_xref="GI:23344063"
                /translation="MPQLSIETFVSQYFWFLVILIVFYLFTISNLVPRIAEIKKTRKK
                CMDEESIVTISSDIQTLNNFNPAKQINVINNSVSNNHINTIVSSFENANKNWIKK"
gene            13729. .14760
                /gene="nad1"
CDS             13729. .14760
                /gene="nad1"
                /EC_number="1.6.5.3"
                /codon_start=1
                /transl_table=4
                /product="NADH dehydrogenase subunit 11"
                /protein_id="AAN28340.1"
                /db_xref="GI:23344064"
                /translation="MTLNLILTIIEIILKILAIVVPVLIAVAYLTLAERKVLASMQLR
                KGPTAVGVFGLLQPLADGLKLFVKEAVLPTHANMIIFVVSPVAAFTLALISWAVIPYN
                EGKVIADVNIGLLYIFAVSSISVYAILMSGWASNSKYAFFGAIRAAAQMISYEVSMGL
                ILLSVILCVGDINITSIVLAQKNIWFIIPLFPAFIMFLVSALAETNRAPFDLTEGESE
                LVSGYNVEYSAIFFTLFFLAEYTHIIFMSILTSLLFFGGWLPPFDFFIFNLIPGEIWL
                ALKTGFIIFVFVWVRATFPRFRYNTLMELMWKSYLPLSLGFLILTASILIAFNDYVQI
                SKFINLLIY"
gene            15143. .15214
                /gene="trnT"
tRNA            15143. .15214
                /gene="trnT"
                /product="tRNA-Thr"
gene            15536. .15617
                /gene="trnS"
tRNA            15536. .15617
                /gene="trnS"
                /product="tRNA-Ser"
gene            15742. .15816
                /gene="trnH"
tRNA            15742. .15816
                /gene="trnH"
                /product="tRNA-His"
gene            16105. .16500
                /gene="rps12"
CDS             16105. .16500
                /gene="rps12"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein S12"
                /protein_id="AAN28341.1"
                /db_xref="GI:23344065"
                /translation="MTTINQLSKKNIRKRKTRITAGSKFLKNQGLINPQLKGVCIRVL
                TRKPKKPNSAQRKIAKVKLTNGLIVDAFIPGEGHSLQEHGIVLIRGGRVKDLPGIKLK
                CVRGKYDLAGVIGRKTSRSKYGTKKPKTK"
gene            17571. .18104
                /gene="rpl16"
CDS             17571. .18104
                /gene="rpl16"
                /codon_start=1
                /transl_table=4
                /product="ribosomal protein L16"
                /protein_id="AAN28342.1"
                /db_xref="GI:23344066"
                /translation="MLNKRPKRTKYSKYQKGRINSNLPVNKSNLFQKNLENKDFRYLL
                IAMEPLRLTASHIAAGELAIKRELNIKFKSSNLSITDIEGNKINIKGDYNLHVFPHLP
                VTKKPAEVRMGKGKGSIDYWMTRIRIGSPIFELKFKNANINNNKNIQIRCFNIAKQIF
                DVVNSKLGIKTKIVKVS"
gene            19351. .19422
                /gene="trnR"
tRNA            19351. .19422
                /gene="trnR"
```

```
                     /product="tRNA-Arg"
     gene            19569. .23450
                     /gene="nad5"
     CDS             join(19569. .20318,22134. .23450)
                     /gene="nad5"
                     /EC_number="1.6.5.3"
                     /codon_start=1
                     /transl_table=4
                     /product="NADH dehydrogenase subunit 5"
                     /protein_id="AAN28343.1"
                     /db_xref="GI:23344067"
                     /translation="MYYLILFLPLLSAVIAGLGGRFIGTKGAGIITTICIMITSLFSW
                     IAFYEVGINMSPVYIPLWKWINVGTYLDMSLGLMYDSLTVTMLILVTTVSSLVHLYSI
                     SYMDGDPHLPRFMSYLSLFTFSMVMLVTADSYLLCVWGWESVGICSYLLINFWYTRIQ
                     ANKSAIKAMVMNRVGDMTLFLGMFAIFVVFNTLDYSTIFSSTYKLSDYTIIFFNYEIH
                     ALTLISILLFLGAVGKSAQLGLHTWLPDAMEGPTPVSALIHAATMVTAGVFLLIRSSP
                     IIEYAPKALVIITIVGTLTALFAASIGLVQNDIKKVIAYSTCSQLGYMVLACGLSNYN
                     VALFHLFTHGFFKALLFLGAGSVIHAMGDEQDMRKLGGLIKRIPYTYSMMLIGSLSLM
                     GFPFLAGFYSKDAILELAYGKYTIDGTFAHLLGTLAAGMTAYYSIRLIYMTFLSDTNG
```

Query Match 4.6%; Score 56.4; DB 3; Length 76568;
Best Local Similarity 43.2%; Pred. No. 1.5;
Matches 364; Conservative 0; Mismatches 476; Indels 2; Gaps 2;

```
Qy       367 AAATTGGTTTATTCTTAACTACCTTTACTTTATTTGTCACTGGATGTTTGACTCAAAAGT 426
Db     64637 AATTTTTATTAATTTTTATTAATTTTTATTAATTTTTATTAATTTTTATTAATTTTTATT 64578

Qy       427 GGTATTTCTCTATTCCCCTTCTTGTCTTAATGCAAATCATCAGTGGTTGGATTGGTCACT 486
Db     64577 AATTTTTATTAATTTTTATTAATTTTTATTAATTTATTAAACCTATCAAAATTCTATATT 64518

Qy       487 CTATGAACCACAATCGTAACCCTATATTAAGAAAATTCGCTTTAGTCTACGCTCCTCTTT 546
Db     64517 CTATTATAAACTACAAATATTTATATTTATAAAATTTAAATTTTATATTTTATTATTATT 64458

Qy       547 GTGGTGGTTTCTCTAATAAATGGTGGGGTAGGAAGCACAATCAACATCATATGTTCACAA 606
Db     64457 TTTATTAATTTTATTAATTTTTATTAATTTATTAAACCTATCAAAATTCTATATTCTATT 64398

Qy       607 ACAACATTCTAAAGGACGAAGATATCTAACACGATTACAAATTGTGGTAATTCCCCTTCT 666
Db     64397 ATAAACTACAAATATTTATATTTATATTTAATTTTATAAAATTTAAATTTTTATTTTATT 64338

Qy       667 TATTTTTAAAGTGGAAATTAGACTCCATCTTAGCTTCTTATTATGAATTTGAAGGAATCT 726
Db     64337 TTATTTATTATATTAAATATTAAAAATTATTTTTATTAGTTATATACCTAATATGAATTT 64278

Qy       727 TCCTTGCCTTGCACTGGGTATTATTATTCAACTAAAACTTCTATATCGTAATTCTTTCTG 786
Db     64277 TGAAATAATATATTTAATAATAATAATTTAAAGATATTTTCAATATTTTTTATAAAATTA 64218

Qy       787 AATTGATTGCTGGTTTCTTCAGTGCTTCTATTCTTGTTGGAAATCATGAAAATGAAATGA 846
Db     64217 ATATTTTTGATAATTTTTATAAAATTAATATTTTTGATAATTTTTATAAAATTAATATTT 64158

Qy       847 AATTCGAAAGAAGAATCACTTTACCATTTTTCGAACATCAAATAGCTGCAAGCAGAAACT 906
Db     64157 TAGATAATTTTATATAATTTTATATAATTTTATATAATATTTAACATAAATAAAATAATA 64098

Qy       907 ACGCTTTCCACGACATATTCTCTCTACTTATTATGGGTGGTATGTAATATTAGACTGAAC 966
Db     64097 TTTAATTTTTATTTTTAATTTTTTTTATAAATTTATTTAATAATTATGATTCAAATGGAG 64038

Qy       967 ATCACTTTTTCCCATAAATTCCTTTCTACAGATTACCCAAAGCTCGTGTCATAATTGCTG 1026
Db     64037 CATA-TTATTTTATTAATTATATATCTATTTAATATTAAACATTATTATCCTTTTTAATT 63979

Qy      1027 AAGAATTAAAGAAGTGGAACCTTAAGATTCATGAAGGACCTATTTTTGAAAAATCTCACC 1086
Db     63978 TATACATTATTTGTATTAAACATTATTTTATTAAAAATTATTTTATTTAAAATTATTTTA 63919

Qy      1087 TTTGAAAATAAATAAATTTATTTTAAATGCATATTT-TATTAGTAATACTAACAATTGTA 1145
```

```
Db     63918 TTTTTTATAATTTTATTTTTTATAATTTTATTTTTTATAATTTTATTTTTTATAATTTTA 63859

Qy      1146 GGAAATGTGTTATGGTTTGTTTACTTATTACTTTTTAATCTGAGAAAACAGTCTTAACAA 1205
Db     63858 TTTTATTTTTTATTTTTATTAAACATTTTCCTATTAAAATTTGGATATTAAAGATATTAA 63799

Qy      1206 AA 1207
Db     63798 AA 63797
```

Search completed: January 19, 2005, 15:43:22
Job time : 5537 secs